

CF015501 US / mao

日本国特許庁

JAPAN PATENT OFFICE



09/891,389

Hiroyuki Kimura
July 27, 2001

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日

Date of Application:

2000年 7月14日

出願番号

Application Number:

特願2000-214140

出願人

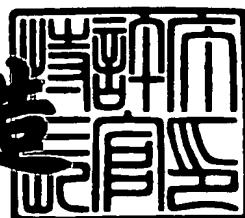
Applicant(s):

キヤノン株式会社

2001年 7月19日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



出証番号 出証特2001-3064556

【書類名】 特許願

【整理番号】 4176056

【提出日】 平成12年 7月14日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 3/00

【発明の名称】 音声出力装置、音声出力システム、音声出力方法及び記憶媒体

【請求項の数】 48

【発明者】

【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

【氏名】 後藤 裕典

【発明者】

【住所又は居所】 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

【氏名】 木村 裕行

【特許出願人】

【識別番号】 000001007

【氏名又は名称】 キヤノン株式会社

【代表者】 御手洗 富士夫

【代理人】

【識別番号】 100081880

【弁理士】

【氏名又は名称】 渡部 敏彦

【電話番号】 03(3580)8464

【手数料の表示】

【予納台帳番号】 007065

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9703713

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声出力装置、音声出力システム、音声出力方法及び記憶媒体

【特許請求の範囲】

【請求項1】 テキストデータを合成音声に変換して出力する音声出力装置であって、

前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力手段とを有することを特徴とする音声出力装置。

【請求項2】 前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする請求項1記載の音声出力装置。

【請求項3】 前記音声出力手段は、前記異なる種類の音声に対応する複数種類の音声素片データを記憶する音声素片記憶部と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成部とを有することを特徴とする請求項1記載の音声出力装置。

【請求項4】 前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする請求項3記載の音声出力装置。

【請求項5】 前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする請求項1記載の音声出力装置。

【請求項6】 所定数の複数種類の音声のうち、任意の音声を選択する選択手段を備え、前記音声出力手段は前記選択された音声に従い合成音声を生成して出力することを特徴とする請求項1記載の音声出力装置。

【請求項7】 前記異なる種類の音声は互いに高さが異なることを特徴とする請求項1記載の音声出力装置。

【請求項8】 前記音声出力手段は、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする請求項1記載の音声出力装置。

【請求項9】 前記異なる種類の音声は、互いに異なる年齢に対応する音声であることを特徴とする請求項1記載の音声出力装置。

【請求項10】 テキストデータを合成音声に変換して出力する音声出力裝

置であって、

前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させる音声出力手段とを有することを特徴とする音声出力装置。

【請求項11】 使用する前記発声手段を任意に設定可能な設定手段を有することを特徴とする請求項10記載の音声出力装置。

【請求項12】 インターネットを介して前記テキストデータで会話を行うシステムに適用可能であることを特徴とする請求項1乃至11の何れかに記載の音声出力装置。

【請求項13】 テキストデータを合成音声に変換して出力する音声出力装置と、該音声出力装置に前記テキストデータを送信する外部装置とを具備してなる音声出力システムであって、

前記音声出力装置は、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力手段とを有することを特徴とする音声出力システム。

【請求項14】 前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする請求項13記載の音声出力システム。

【請求項15】 前記音声出力手段は、前記異なる種類の音声に対応する複数種類の音声素片データを記憶する音声素片記憶部と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成部とを有することを特徴とする請求項13記載の音声出力システム。

【請求項16】 前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする請求項15記載の音声出力システム。

【請求項17】 前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする請求項13記載の音声出力システム。

【請求項18】 前記音声出力装置は、所定数の複数種類の音声のうち、任

意の音声を選択する選択手段を備え、前記音声出力手段は前記選択された音声に従い合成音声を生成して出力することを特徴とする請求項13記載の音声出力システム。

【請求項19】 前記異なる種類の音声は互いに高さが異なることを特徴とする請求項13記載の音声出力システム。

【請求項20】 前記音声出力手段は、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする請求項13記載の音声出力システム。

【請求項21】 前記異なる種類の音声は、互いに異なる年齢に対応する音声であることを特徴とする請求項13記載の音声出力システム。

【請求項22】 テキストデータを合成音声に変換して出力する音声出力装置と、該音声出力装置に前記テキストデータを送信する外部装置とを具備してなる音声出力システムであって、

前記音声出力装置は、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させる音声出力手段とを有することを特徴とする音声出力システム。

【請求項23】 前記音声出力装置は、使用する前記発声手段を任意に設定可能な設定手段を有することを特徴とする請求項22記載の音声出力システム。

【請求項24】 インターネットを介して前記テキストデータで会話を行うシステムに適用可能であることを特徴とする請求項13乃至23の何れかに記載の音声出力システム。

【請求項25】 テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法であって、

前記テキストデータの音声波形を生成する音声波形生成工程と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力工程とを有することを特徴とする音声出力方法。

【請求項26】 前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする請求項25記載の音声出力方法。

【請求項27】 前記音声出力工程は、前記異なる種類の音声に対応する複

数種類の音声素片データを記憶する音声素片記憶工程と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成工程とを有することを特徴とする請求項25記載の音声出力方法。

【請求項28】 前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする請求項27記載の音声出力方法。

【請求項29】 前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする請求項25記載の音声出力方法。

【請求項30】 所定数の複数種類の音声のうち、任意の音声を選択する選択工程を有し、前記音声出力工程では前記選択された音声に従い合成音声を生成して出力することを特徴とする請求項25記載の音声出力方法。

【請求項31】 前記異なる種類の音声は互いに高さが異なることを特徴とする請求項25記載の音声出力方法。

【請求項32】 前記音声出力工程では、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする請求項25記載の音声出力方法。

【請求項33】 前記異なる種類の音声は、互いに異なる年齢に対応する音声であることを特徴とする請求項25記載の音声出力方法。

【請求項34】 テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法であって、

前記テキストデータの音声波形を生成する音声波形生成工程と、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させる音声出力工程とを有することを特徴とする音声出力方法。

【請求項35】 使用する前記発声手段を任意に設定可能な設定工程を有することを特徴とする請求項34記載の音声出力方法。

【請求項36】 インターネットを介して前記テキストデータで会話をを行うシステムに適用可能であることを特徴とする請求項25乃至35の何れかに記載の音声出力方法。

【請求項37】 テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法を実行するプログラムを記憶したコンピュータにより読み出し可能な記憶媒体であって、

前記音声出力方法は、前記テキストデータの音声波形を生成するように制御する音声波形生成ステップと、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させるように制御する音声出力ステップとを有することを特徴とする記憶媒体。

【請求項38】 前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする請求項37記載の記憶媒体。

【請求項39】 前記音声出力ステップは、前記異なる種類の音声に対応する複数種類の音声素片データを記憶するように制御する音声素片記憶ステップと、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成するように制御する音声波形生成ステップとを有することを特徴とする請求項37記載の記憶媒体。

【請求項40】 前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする請求項39記載の記憶媒体。

【請求項41】 前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする請求項37記載の記憶媒体。

【請求項42】 所定数の複数種類の音声のうち、任意の音声を選択する選択ステップを有し、前記音声出力ステップでは前記選択された音声に従い合成音声を生成して出力するように制御することを特徴とする請求項37記載の記憶媒体。

【請求項43】 前記異なる種類の音声は互いに高さが異なることを特徴とする請求項37記載の記憶媒体。

【請求項44】 前記音声出力ステップでは、所定数の複数種類の音声を所定の順序で選択的に出力するように制御することを特徴とする請求項37記載の記憶媒体。

【請求項45】 前記異なる種類の音声は、互いに異なる年齢に対応する音

声であることを特徴とする請求項37記載の記憶媒体。

【請求項46】 テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法を実行するプログラムを記憶したコンピュータにより読み出し可能な記憶媒体であって、

前記音声出力方法は、前記テキストデータの音声波形を生成するように制御する音声波形生成ステップと、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させるように制御する音声出力ステップとを有することを特徴とする記憶媒体。

【請求項47】 使用する前記発声手段を任意に設定可能な設定ステップを有することを特徴とする請求項46記載の記憶媒体。

【請求項48】 インターネットを介して前記テキストデータで会話をを行うシステムに適用可能であることを特徴とする請求項37乃至47の何れかに記載の記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、音声出力装置、音声出力システム、音声出力方法及び記憶媒体に関し、特に、テキストデータを合成音声に変換して出力する場合に好適な音声出力装置、音声出力システム、音声出力方法及び記憶媒体に関する。

【0002】

【従来の技術】

従来、文字情報を音声出力する機能を有する音声合成装置がある。従来の音声合成装置において、音声出力するデータは、予め電子化されたテキストデータとして準備されたものでなければならなかった。即ち、そのテキストデータは、パソコン用コンピュータ上におけるエディタやワードプロセッサなどで作成した文章や、インターネット上のHTML(HyperText Markup Language)文章である。

【0003】

また、音声合成装置から上記のようなテキストデータを音声で出力する場合、

1入力からのテキストデータを、音声合成装置に予め設定されている1種類の音声で出力していた。

【0004】

【発明が解決しようとする課題】

しかしながら、上記従来の音声合成装置においては、同時に複数のテキストデータの入力を受け付け、更にそれらの合成音声出力を重ね合わせて出力し、それらを聞き分けられるように出力することができないという問題があった。

【0005】

本発明は、上述した点に鑑みなされたものであり、複数のテキストデータの合成音声を重ね合わせて発声する場合、複数のテキストデータをそれぞれ異なる種類の音声で音声合成し出力させることで、容易に複数のテキストデータの音声を聞き分けることができる音声出力装置、音声出力システム、音声出力方法及び記憶媒体を提供することを第一の目的とする。

【0006】

また、本発明は、複数のテキストデータの合成音声を重ね合わせて発声する場合、複数のテキストデータの音声をそれぞれ異なる発声手段で発声させることで、容易に複数のテキストデータの音声を聞き分けることができる音声出力装置、音声出力システム、音声出力方法及び記憶媒体を提供することを第二の目的とする。

【0007】

【課題を解決するための手段】

上記目的を達成するため、請求項1記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置であって、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力手段とを有することを特徴とする。

【0008】

上記目的を達成するため、請求項2記載の発明は、前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする。

【0009】

上記目的を達成するため、請求項3記載の発明は、前記音声出力手段は、前記異なる種類の音声に対応する複数種類の音声素片データを記憶する音声素片記憶部と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成部とを有することを特徴とする。

【0010】

上記目的を達成するため、請求項4記載の発明は、前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする。

【0011】

上記目的を達成するため、請求項5記載の発明は、前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする。

【0012】

上記目的を達成するため、請求項6記載の発明は、所定数の複数種類の音声のうち、任意の音声を選択する選択手段を備え、前記音声出力手段は前記選択された音声に従い合成音声を生成して出力することを特徴とする。

【0013】

上記目的を達成するため、請求項7記載の発明は、前記異なる種類の音声は互いに高さが異なることを特徴とする。

【0014】

上記目的を達成するため、請求項8記載の発明は、前記音声出力手段は、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする。

【0015】

上記目的を達成するため、請求項9記載の発明は、前記異なる種類の音声は、互いに異なる年齢に対応する音声であることを特徴とする。

【0016】

上記目的を達成するため、請求項10記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置であって、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータの音声出力が重なり合

うことを検知した場合は、それぞれの音声を異なる発声手段から出力させる音声出力手段とを有することを特徴とする。

【0017】

上記目的を達成するため、請求項11記載の発明は、使用する前記発声手段を任意に設定可能な設定手段を有することを特徴とする。

【0018】

上記目的を達成するため、請求項12記載の発明は、インターネットを介して前記テキストデータで会話を行うシステムに適用可能であることを特徴とする。

【0019】

上記目的を達成するため、請求項13記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置と、該音声出力装置に前記テキストデータを送信する外部装置とを具備してなる音声出力システムであって、前記音声出力装置は、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力手段とを有することを特徴とする。

【0020】

上記目的を達成するため、請求項14記載の発明は、前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする。

【0021】

上記目的を達成するため、請求項15記載の発明は、前記音声出力手段は、前記異なる種類の音声に対応する複数種類の音声素片データを記憶する音声素片記憶部と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成部とを有することを特徴とする。

【0022】

上記目的を達成するため、請求項16記載の発明は、前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする。

【0023】

上記目的を達成するため、請求項17記載の発明は、前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする。

【0024】

上記目的を達成するため、請求項18記載の発明は、前記音声出力装置は、所定数の複数種類の音声のうち、任意の音声を選択する選択手段を備え、前記音声出力手段は前記選択された音声に従い合成音声を生成して出力することを特徴とする。

【0025】

上記目的を達成するため、請求項19記載の発明は、前記異なる種類の音声は互いに高さが異なることを特徴とする。

【0026】

上記目的を達成するため、請求項20記載の発明は、前記音声出力手段は、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする。

【0027】

上記目的を達成するため、請求項21記載の発明は、前記異なる種類の音声は互いに異なる年齢に対応する音声であることを特徴とする。

【0028】

上記目的を達成するため、請求項22記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置と、該音声出力装置に前記テキストデータを送信する外部装置とを具備してなる音声出力システムであって、前記音声出力装置は、前記テキストデータの音声波形を生成する音声波形生成手段と、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させる音声出力手段とを有することを特徴とする。

【0029】

上記目的を達成するため、請求項23記載の発明は、前記音声出力装置は、使用する前記発声手段を任意に設定可能な設定手段を有することを特徴とする。

【0030】

上記目的を達成するため、請求項24記載の発明は、インターネットを介して前記テキストデータで会話をを行うシステムに適用可能であることを特徴とする。

【0031】

上記目的を達成するため、請求項25記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法であって、前記テキストデータの音声波形を生成する音声波形生成工程と、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させる音声出力工程とを有することを特徴とする。

【0032】

上記目的を達成するため、請求項26記載の発明は、前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする。

【0033】

上記目的を達成するため、請求項27記載の発明は、前記音声出力工程は、前記異なる種類の音声に対応する複数種類の音声素片データを記憶する音声素片記憶工程と、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成する音声波形生成工程とを有することを特徴とする。

【0034】

上記目的を達成するため、請求項28記載の発明は、前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする。

【0035】

上記目的を達成するため、請求項29記載の発明は、前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする。

【0036】

上記目的を達成するため、請求項30記載の発明は、所定数の複数種類の音声のうち、任意の音声を選択する選択工程を有し、前記音声出力工程では前記選択された音声に従い合成音声を生成して出力することを特徴とする。

【0037】

上記目的を達成するため、請求項31記載の発明は、前記異なる種類の音声は互いに高さが異なることを特徴とする。

【0038】

上記目的を達成するため、請求項32記載の発明は、前記音声出力工程では、所定数の複数種類の音声を所定の順序で選択的に出力することを特徴とする。

【0039】

上記目的を達成するため、請求項33記載の発明は、前記異なる種類の音声は、互いに異なる年齢に対応する音声であることを特徴とする。

【0040】

上記目的を達成するため、請求項34記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法であって、前記テキストデータの音声波形を生成する音声波形生成工程と、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる发声手段から出力させる音声出力工程とを有することを特徴とする。

【0041】

上記目的を達成するため、請求項35記載の発明は、使用する前記発声手段を任意に設定可能な設定工程を有することを特徴とする。

【0042】

上記目的を達成するため、請求項36記載の発明は、インターネットを介して前記テキストデータで会話を行うシステムに適用可能であることを特徴とする。

【0043】

上記目的を達成するため、請求項37記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法を実行するプログラムを記憶したコンピュータにより読み出し可能な記憶媒体であって、前記音声出力方法は、前記テキストデータの音声波形を生成するように制御する音声波形生成ステップと、複数の前記テキストデータをそれぞれ異なる種類の音声で音声合成し出力させるように制御する音声出力ステップとを有することを特徴とする。

【0044】

上記目的を達成するため、請求項38記載の発明は、前記異なる種類の音声は互いに周波数帯域が異なることを特徴とする。

【0045】

上記目的を達成するため、請求項39記載の発明は、前記音声出力ステップは、前記異なる種類の音声に対応する複数種類の音声素片データを記憶するように制御する音声素片記憶ステップと、前記異なる種類の音声に対応する処理パラメータに従って前記音声素片データを処理し、合成音声を生成するように制御する音声波形生成ステップとを有することを特徴とする。

【0046】

上記目的を達成するため、請求項40記載の発明は、前記処理パラメータは、周波数帯域と、音声レベルと、音声速度のうちの少なくとも1つを含むことを特徴とする。

【0047】

上記目的を達成するため、請求項41記載の発明は、前記異なる種類の音声は、互いに異なる性別に対応する音声であることを特徴とする。

【0048】

上記目的を達成するため、請求項42記載の発明は、所定数の複数種類の音声のうち、任意の音声を選択する選択ステップを有し、前記音声出力ステップでは前記選択された音声に従い合成音声を生成して出力するように制御することを特徴とする。

【0049】

上記目的を達成するため、請求項43記載の発明は、前記異なる種類の音声は互いに高さが異なることを特徴とする。

【0050】

上記目的を達成するため、請求項44記載の発明は、前記音声出力ステップでは、所定数の複数種類の音声を所定の順序で選択的に出力するように制御することを特徴とする。

【0051】

上記目的を達成するため、請求項45記載の発明は、前記異なる種類の音声は互いに異なる年齢に対応する音声であることを特徴とする。

【0052】

上記目的を達成するため、請求項4 6記載の発明は、テキストデータを合成音声に変換して出力する音声出力装置に適用される音声出力方法を実行するプログラムを記憶したコンピュータにより読み出し可能な記憶媒体であって、前記音声出力方法は、前記テキストデータの音声波形を生成するように制御する音声波形生成ステップと、複数の前記テキストデータの音声出力が重なり合うことを検知した場合は、それぞれの音声を異なる発声手段から出力させるように制御する音声出力ステップとを有することを特徴とする。

【0053】

上記目的を達成するため、請求項4 7記載の発明は、使用する前記発声手段を任意に設定可能な設定ステップを有することを特徴とする。

【0054】

上記目的を達成するため、請求項4 8記載の発明は、インターネットを介して前記テキストデータで会話をを行うシステムに適用可能であることを特徴とする。

【0055】

【発明の実施の形態】

以下、本発明の実施の形態を図面に基づいて詳細に説明する。

【0056】

【第1の実施の形態】

本発明の第1の実施の形態は、他のコンピュータ（サーバコンピュータ）から非同期に送られてくるテキストデータを音声出力するシステムにおいて、1つのテキストデータの音声出力が終了する前に、次のテキストデータが送られてきたとき、先に音声出力中の性別の音声とは別の性別の音声で次のテキストデータを読み上げるシステムである。

【0057】

本実施形態では、音声出力に重なりがないときに、通常の性別として使用する性別を主性別と呼び、音声が重なった場合に、次のテキストデータを読み上げるために使用される、先に音声出力中の主性別とは異なる性別を副性別と呼ぶ（図5参照）。但し、副性別で音声出力中に次のテキストデータの音声出力をを行うときは主性別で行う。

【0058】

図1は本発明の第1の実施の形態に係る音声合成装置の構成例を示すブロック図である。本発明の第1の実施の形態に係る音声合成装置は、C P U 1 0 1、ハードディスクコントローラ（H D C）1 0 2、プログラム1 1 3・辞書1 1 4・音声素片データ1 1 5を有するハードディスク（H D）1 0 3、キーボード1 0 4、ポインティングデバイス（P D）1 0 5、R A M 1 0 6、通信回線インターフェイス（I / F）1 0 7、V R A M 1 0 8、表示コントローラ1 0 9、モニタ1 1 0、サウンドカード1 1 1、スピーカ1 1 2、描画部1 1 6を備えている。図中1 5 0はサーバコンピュータである。

【0059】

上記各部の構成を詳述すると、C P U 1 0 1は、本装置全体の制御を行う中央演算処理装置であり、後述の図4のフローチャートに示す処理を実行する。ハードディスクコントローラ1 0 2は、ハードディスク1 0 3内のデータやプログラムの制御を行う。ハードディスク1 0 3内には、プログラム1 1 3、音声波形生成部（後述）において入力された漢字かな混じり文を解析し読み情報を得る時に参照される漢字等の読みやアクセント情報が登録されている辞書1 1 4、発声する文字列に従って音声の素片を接続する際に必要になる音声素片データ1 1 5が格納されている。この音声素片データ1 1 5には、少なくとも男性の声の出力となる音声素片データと、女性の声の出力となる音声素片データの2種類の音声素片データが含まれている。これら2種類の音声素片データは、性別に応じてその基本周波数が異なっている。

【0060】

キーボード1 0 4は、文字・数字・記号等の入力に用いる。ポインティングデバイス1 0 5は、プログラム開始などの指示に用いるものであり、例えばマウスやディジタイザなどから構成されている。R A M 1 0 6は、プログラムやデータを格納する。通信回線インターフェイス1 0 7は、外部のサーバコンピュータ1 5 0との間でデータのやり取りを行う。本実施形態では通信形態はT C P / I P（Transmission Control Protocol/Internet Protocol）を用いている。表示コントローラ1 0 9は、V R A M 1 0 8に格納された映像データを映像信号として

モニタ110に出力する制御を行う。サウンドカード111は、CPU101で生成されRAM106に格納された音声波形データをスピーカ112を通して出力している。116は描画部であり、CPU101の制御の下で、RAM106等を用いてモニタ110への表示画像データを生成する。

【0061】

図2は本発明の第1の実施の形態に係る音声合成装置のプログラムのモジュール関係を示す説明図である。本発明の第1の実施の形態に係る音声合成装置は、辞書114、音声素片データ115、メインルーチン初期化部201、音声処理初期化部202、通信データ処理部204、通信データ格納部206、表示文章データ格納部207、文章表示部208、音声波形生成部209（音声波形生成手段）、音声出力部210（音声出力手段）、初期化部203・受信部205を有する通信処理部211、音響パラメータ212、出力パラメータ213を備えている。

【0062】

上記各部の機能を詳述すると、本実施形態のシステムが起動すると、先ず、メインルーチン220のメインルーチン初期化部201にてプログラム全体の初期化が行われる。次に、通信処理部211の初期化部203にて通信部230の初期化が行われ、音声処理初期化部202にて音声部240の初期化が行われる。本実施形態では通信形態はTCP/IPを用いている。

【0063】

通信処理部211の初期化部203にて通信部230の初期化が終了すると、通信処理部211の受信部205が起動し、サーバコンピュータ150から音声合成装置に送信されるテキストデータを受信できるようになる。通信処理部211の受信部205にてテキストデータを受信すると、受信されたテキストデータは通信データ格納部206に格納される。

【0064】

メインルーチン初期化部201にてメインルーチン220全体の初期化が終了すると、通信データ処理部204は通信データ格納部206の監視を開始する。受信されたテキストデータが通信データ格納部206に格納されると、通信データ

タ処理部204はそのテキストデータを読み込み、モニタ110に表示するための表示文章を格納する表示文章データ格納部207にテキストデータを格納する。

【0065】

文章表示部208は、表示文章データ格納部207にデータがあるのを検知すると、そのデータをモニタ110に表示できる形態に変換して、VRAM108上に置く。この結果、モニタ110に表示文章が表示されることになる。この際、テキストデータに何らかの処理を施して表示文章とする場合は、その処理を通信データ処理部204にて行う。

【0066】

また、通信データ処理部204は、受信されたテキストデータを音声波形生成部209に送り、音声波形生成部209にてテキストデータの音声波形の生成を行う。その際、テキストデータに何らかの処理を施して音声波形を生成する場合は、その処理を通信データ処理部204にて行う。音声波形生成部209にて辞書114、音声素片データ115、音響パラメータ212を参照しながら、受信したテキストデータの音声波形を生成する。その処理の流れは図4を使って後で説明する。生成された波形はミキシング機能を持つ音声出力部210に渡される。

【0067】

図3は本発明の第1の実施の形態に係る音声合成装置の音声出力部210の詳細構成を示す説明図である。本発明の第1の実施の形態に係る音声合成装置の音声出力部210は、一時蓄積部601、コントロール部602、音声再生部604、ミキシング部605を備えている。図中603は音声波形である。

【0068】

上記各部の機能を詳述すると、一時蓄積部601は、音声波形生成部209より送られてきた音声波形603を一時的に蓄積する。コントロール部602は、音声出力部210の全体をコントロールするものであり、常時、一時蓄積部601に音声波形603が送られてきたかどうかをチェックしており、音声波形603が送られてきた場合は、音声再生部604に送って音声再生を開始する。

【0069】

音声再生部604は、上記図2の出力パラメータ213から、音声出力に必要な予め設定されたパラメータ（サンプリングレート、データのビット数等のこと）に従って、音声波形603の再生を実行する。

【0070】

尚、音声再生部604は少なくとも2つ以上存在し、コントロール部602は、音声波形603が送られてきた場合は、その時点で使用されていない音声再生部604に音声波形603を送って、再生を実行する。また、音声再生部604をソフト的なプロセスとして構成し、コントロール部602は、音声波形603が送られてくる毎に音声再生部604のプロセスを生成し、音声波形603の再生が終わった時点で、その音声再生部604のプロセスを消滅させるような構成であっても構わない。

【0071】

音声再生部604が出力する個々の音声データは、少なくとも2つ以上の入力部を持つミキシング部605に送られ、ミキシング部605は音声データを合成して、最終的な合成音声データを上記図1のスピーカ112から出力する。この際、コントロール部602は、ミキシング部605に送られる音声データの数に応じて、ミキシングのレベル調整をミキシング部605に対して行う。

【0072】

また、コントロール部602は、音声波形生成部209から、音声出力中かどうかの問い合わせを受け、音声再生部604やミキシング部605の動作状況を調べて、音声波形生成部209に結果を返す機能も持っている。更に、コントロール部602は、音声波形生成部209から、どの性別で出力中かどうかの問い合わせを受け、音声再生部604で再生中の音声波形のデータを調べて、音声波形生成部209に結果を返す機能も持っている。

【0073】

図22は本発明の音声出力方法を実行するプログラム及び関連データが記憶媒体から装置に供給される概念例を示す説明図である。本発明の音声出力方法を実行するプログラム及び関連データは、フロッピディスクやCD-ROM等の記憶

媒体2201を装置2202に装備された記憶媒体ドライブ挿入口2203に挿入することで供給される。その後、プログラム及び関連データを記憶媒体2201から一旦ハードディスクにインストールしハードディスクからRAMにロードするか、或いはハードディスクにインストールせずに直接RAMにロードすることで、プログラム及び関連データを実行することが可能となる。

【0074】

この場合、本発明の第1～第4の実施の形態に係る音声合成装置において本発明の音声出力方法を実行するプログラムを実行する場合は、上記図22を参照して説明したような手順で音声合成装置にプログラム及び関連データを供給するか、或いは音声合成装置に予めプログラム及び関連データを格納しておくことで、プログラム実行が可能となる。

【0075】

図21は本発明の音声出力方法を実行するプログラム及び関連データを記憶した記憶媒体の記憶内容の構成例を示す説明図である。記憶媒体は、例えばボリューム情報2101、ディレクトリ情報2102、プログラム実行ファイル2103（図1や後述の図7の、プログラム113に相当）、プログラム関連データファイル2104（図1や後述の図7の、辞書114、音声素片データ115等に相当）等の記憶内容で構成される。本発明の音声出力方法を実行するプログラムは、後述の図4（第1の実施の形態）、図10（第2の実施の形態）、図13・図14（第3の実施の形態）、図17・図18（第4の実施の形態）の各フローチャートに基づきプログラムコード化されたものである。

【0076】

次に、上記の如く構成された本発明の第1の実施の形態に係る音声合成装置の動作を図4、図6を参照しながら詳細に説明する。下記の処理は上記図1に示したCPU101の制御のもとに実行される。

【0077】

図4は音声合成装置の通信データ処理部204から音声波形生成部209に送られてきたテキストデータを音声出力する処理を示す流れ図である。先ずステップS401で、音声出力部210のコントロール部602に対して現在音声を出

力中か問い合わせる。その結果、音声を出力中でなければ、ステップS408にて、音声の性別を主性別（例えば男性）に設定し、ステップS404に進む。

【0078】

上記ステップS401で現在音声を出力中であれば、ステップS402にて、音声出力部210のコントロール部602に対して現在出力中の音声が主性別か副性別かを問い合わせ、現在出力中の音声が主性別（例えば男性）ならば、ステップS403にて、音声の性別を副性別（例えば女性）に設定する。上記ステップS402で現在出力中の音声が副性別（例えば女性）ならば、ステップS408にて、音声の性別を主性別（例えば男性）に設定する。

【0079】

ステップS404にて、上記ステップS403もしくはステップS408で切り替えた音声の性別の情報に合わせて、音声素片データ115の中から適切な性別の音声素片データが選択される。ステップS405にて、辞書114を使ってテキストデータの言語解析を行い、テキストデータの読みと音調成分を生成する。更にステップS406にて、音響パラメータ212に収納された予め設定された、声の高さ（周波数帯域）・アクセント（音声レベル）・発声速度等に関するパラメータのうち、ステップS403もしくはS408にて選択された性別に応じたパラメータに従って、上記ステップS404にて選択されていた音声素片データと、上記ステップS405にて解析されたテキストデータの読みと音調成分を使って音声波形を生成する。即ち、主性別が選択された場合には主性別に対応したパラメータに従って音声波形を生成し、副性別が選択された場合には副性別に対応したパラメータに従って音声波形を生成する。

【0080】

ステップS407にて、上記ステップS406で生成した音声波形を音声出力部210に渡して音声出力を行う。音声波形が音声出力部210に送られると、音声再生部604の1つを使って音声の再生が行われるが、音声再生部604にて現在再生中の音声がある場合は、ミキシング部605にて現在再生中の音声に新しく渡された音声がミキシングされて音声出力が行われることになる。現在再生中の音声がない場合は、再生された音声はミキシング部605を通過するが、

何の処理もなされずにそのままの音声出力が行われることになる。

【0081】

このように、複数の音声出力が重なり合うことを検知した時、これらの音声を性別の異なる音声で出力することによって、複数の音声が重なっても容易に聞き取れるようになる。

【0082】

図5は音声合成装置における主性別での出力音声と副性別での出力音声の時間的な関係を示す概念図、図6は音声合成装置における主性別を設定する方法を示す説明図である。

【0083】

キーボード104またはP D 1 0 5により、音声出力設定画面の指示があると、C P U 1 0 1は描画部116を用いて図6に示す設定画面の画像データを生成し、表示コントローラ109によりモニタ110上に表示する。

【0084】

そして、ユーザは、P D 1 0 5を用いて、図6の設定画面（設定手段）の503で主性別を男性、女性から選択する。「O K」ボタン501を押下することにより、上記図1のR A M 1 0 6上に格納されている主性別の変数が書き換えられ、選択が完了する。また、「c a n c e l」ボタン502を押下したときは、R A M 1 0 6上に格納されている主性別の変数が書き換えられず、選択は破棄されて性別設定モードは終了する。尚、副性別は自動的に主性別の反対の性別が選択される。

【0085】

以上説明したように、本発明の第1の実施の形態に係る音声合成装置によれば、複数の音声出力の重なり合いを検知し、それぞれの音声を違う性別の音声で出力することで、聞き取りが容易になるという効果を奏する。

【0086】

本実施形態を用いれば、例えばインターネットで接続された複数のユーザ端末同士がサーバコンピュータを介してテキストデータで会話をを行うチャットシステムにおいて、サーバコンピュータから送られてきた他のユーザの発言であるテキ

ストデータを音声出力する際に、複数のユーザからのテキストデータの音声出力が重なり合う場合に聞き取りを容易にすることができるという効果を奏する。

【0087】

[第2の実施の形態]

本発明の第2の実施の形態は、他のコンピュータ（サーバコンピュータ）から非同期に送られてくるテキストデータを音声出力するシステムにおいて、1つのテキストデータの音声出力が終了する前に、次のテキストデータが送られてきたとき、先に出力中の合成音声と、次の合成音声の出力を、それぞれ異なったスピーカで再生させるシステムである。

【0088】

即ち、音声出力に重なりがないときは、通常、コンピュータに接続されている2つのステレオスピーカの両方を使って出力し（2つのスピーカ両方で同じ音声が再生される）、音声が重なった場合には、それぞれの音声が2つのスピーカの片方を使って出力する（片方のスピーカから最初の音声が、もう片方のスピーカから次の音声が再生される）（図11参照）。尚、本実施形態では、2つ以上の音声が重ならない前提で考えているが、3つ以上のスピーカを個別に再生できるシステムであれば、第3の音声、第4の音声等が重なっても対応することが可能である。

【0089】

図7は本発明の第2の実施の形態に係る音声合成装置の概略構成を示すブロック図である。本発明の第2の実施の形態に係る音声合成装置は、CPU101、ハードディスクコントローラ（HDC）102、プログラム113・辞書114・音声素片データ115を有するハードディスク（HD）103、キーボード104、ポインティングデバイス（PD）105、RAM106、通信回線インターフェイス（I/F）107、VRAM108、表示コントローラ109、モニタ110、サウンドカード111、右用スピーカ112R・左用スピーカ112Lを有するスピーカ112（発声手段）、描画部116を備えている。

【0090】

第2の実施の形態における上記第1の実施の形態との相異点を説明すると、C

P U 1 0 1 は、後述の図10のフローチャートに示す処理を実行する。サウンドカード111は、C P U 1 0 1 で生成されR A M 1 0 6 に格納された音声波形データを、スピーカ112（右用スピーカ112R、左用スピーカ112L）を通して出力している。尚、音声合成装置の他の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0091】

図8は本発明の第2の実施の形態に係る音声合成装置のプログラムのモジュール関係を示す説明図である。本発明の第2の実施の形態に係る音声合成装置のプログラムモジュールは、辞書114、音声素片データ115、メインルーチン初期化部201、音声処理初期化部202、通信データ処理部204、通信データ格納部206、表示文章データ格納部207、文章表示部208、音声波形生成部209（音声波形生成手段）、音声出力部210（音声出力手段）、初期化部203・受信部205を有する通信処理部211、音響パラメータ212、出力パラメータ213を備えている。

【0092】

第2の実施の形態における上記第1の実施の形態との相異点を説明すると、音声波形生成部209から音声出力部210へ入力される信号が1つとなっている点と、音声出力の重なりを検出し、それに対するアクションつまりスピーカの出力を変更する制御はすべて音声出力部210で行われる点である。音声合成装置のプログラムモジュールの他の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0093】

図9は本発明の第2の実施の形態に係る音声合成装置のプログラムのモジュールにおける音声出力部210の詳細構成を示す説明図である。本発明の第2の実施の形態に係る音声合成装置の音声出力部210は、一時蓄積部601、コントロール部602、音声再生部604、ミキシング部605を備えている。

【0094】

第2の実施の形態における上記第1の実施の形態との相異点を説明すると、音声再生部604は2つ存在し、コントロール部602は、音声波形603が送ら

れてきた場合は、その時点で使用されていない音声再生部604に音声波形603を送って、再生を実行する。音声再生部604が出力する個々の音声データは、2つの入力部を持つミキシング部605に送られ、ミキシング部605は音声データを合成して、最終的な合成音声データを上記図7に示したスピーカ112（右用スピーカ112R、左用スピーカ112L）から出力する。

【0095】

この際、ミキシング部605は、スピーカ112の2つのスピーカ112R、112Lに出力する音声をスピーカ毎にコントロールすることが可能であり、コントロール部602は、このスピーカ出力のコントロールをミキシング部605に対して行えるようになっている。音声出力部210の他の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0096】

尚、本システムでは、スピーカを2つ使用しているので、最大2つの音声を同時に再生できるが、3つ以上のスピーカを個別にコントロールできるシステムにおいては、その制御できるスピーカの数まで、音声が重なっても対応することが可能である。

【0097】

次に、上記の如く構成された本発明の第2の実施の形態に係る音声合成装置の動作を図10、図12を参照しながら詳細に説明する。下記の処理は上記図7に示したCPU101の制御のもとに実行される。

【0098】

図10は音声合成装置の音声波形生成部209より音声波形が音声出力部210に送られてきた時から音声を出力するまでの処理を示す流れ図である。先ずステップS1001で、音声出力部210のコントロール部602は音声再生部604の動作状態を調べて、現在音声を出力中かどうかを確認する。その結果、音声を出力中でなければ、ステップS1008で、コントロール部602は、この音声を両方のスピーカ112R、112Lを使って再生することをミキシング部605に指示して、音声の再生を実行する。

【0099】

上記ステップS1001で現在音声1つが出力中であれば、ステップS1002に進んで、コントロール部602は、現在音声再生中の音声を第1のスピーカ（112Rまたは112L）で再生し、次の音声を第2のスピーカ（112Lまたは112R）で再生するように、ミキシング部605に指示して、音声の再生を実行する。尚、上記ステップS1001で既に2つの音声再生が実行されている場合は、再度ステップS1001に戻って、出力中の音声が1つ以下になるまで待つようになっている。

【0100】

上記ステップS1002で2つの音声再生を開始した後は、ステップS1003に進んで、どちらかの音声再生が終了するのを待つ。どちらかの音声再生が終了した時は、ステップS1004で、コントロール部602は、残った再生中の音声を両方のスピーカ112R、112Lを使って再生することをミキシング部605に指示して、音声の再生を実行する。

【0101】

このように、2つの音声出力が重なり合うことを検知した時、それぞれの音声を違うスピーカ112R、112Lで出力することによって、2つの音声が重なっても容易に聞き取れるようになる。

【0102】

尚、3つ以上のスピーカを個別に再生できるシステムであれば、音声出力が重なった条件に応じて、スピーカを割り当てるよう設定すれば、3種類以上の音声が重なっても聞き取ることが可能となる。

【0103】

図11は音声合成装置における両方のスピーカでの再生音声と別々のスピーカでの再生音声の時間的な関係を示す概念図、図12は音声合成装置におけるスピーカの設定を行う方法を示す説明図である。

【0104】

キー ボード104またはPD105により、音声出力設定画面の指示があると、CPU101は描画部116を用いて図12に示す設定画面の画像データを生成し、表示コントローラ109によりモニタ110上に表示する。

【0105】

そして、ユーザは、P D 1 0 5 を用いて、図12の設定画面（設定手段）の503で、音声が重なった時に最初の音声を出力するスピーカを選択し、「OK」ボタン501を押下することにより、上記図1のRAM106上に格納されている最初の音声用のスピーカの設定の変数が書き換えられ、選択が完了する。

【0106】

この時、次の音声を出力するスピーカは、自動的にもう1つのスピーカに設定される。また、「cancel」ボタン502を押下したときは、RAM106上に格納されているスピーカの設定の変数が書き換えられず、選択は破棄されてスピーカ設定モードは終了する。尚、3つ以上のスピーカを設定できる場合は、上記503と同じ形式で次の音声用のスピーカ等が選択できるようになっていればよい。

【0107】

以上説明したように、本発明の第2の実施の形態に係る音声合成装置によれば、2つの音声出力が重なり合うのを検知し、それぞれの音声を別のスピーカ112R、112Lで出力することで、聞き取りが容易になるという効果を奏する。

【0108】

本実施形態を用いれば、例えばインターネットで接続された複数のユーザ端末同士がサーバコンピュータを介してテキストデータで会話をを行うチャットシステムにおいて、サーバコンピュータから送られてきた他のユーザの発言であるテキストデータを音声出力する際に、複数のユーザからのテキストデータの音声出力が重なり合う場合に聞き取りを容易にすることができるという効果を奏する。

【0109】

[第3の実施の形態]

本発明の第3の実施の形態は、他のコンピュータ（サーバコンピュータ）から非同期に送られてくるテキストデータを音声出力するシステムにおいて、1つのテキストデータの音声出力が終了する前に、次のテキストデータが送られてきたとき、先に音声出力中の音声とは別の種類の音声で次のテキストデータを読み上げるシステムである。

【0110】

本実施形態では、音声出力に重なりがないときに、通常に使用される音声を第1音声と呼び、音声が重なった場合に、次のテキストデータを読み上げるために使用される、先に音声出力中の第1音声とは種類が異なる音声を第2音声と呼ぶ（図15参照）。本実施形態では、2つ以上の音声が重ならない前提で考えているが、更に音声が重なることが予想される場合は、第3音声、第4音声と用意しておけばよい。

【0111】

本発明の第3の実施の形態に係る音声合成装置は、上記第1の実施の形態と同様に、CPU101、ハードディスクコントローラ（HDC）102、プログラム113・辞書114・音声素片データ115を有するハードディスク（HD）103、キーボード104、ポインティングデバイス（PD）105、RAM106、通信回線インターフェイス（I/F）107、VRAM108、表示コントローラ109、モニタ110、サウンドカード111、スピーカ112、描画部116を備えている（上記図1参照）。

【0112】

第3の実施の形態における上記第1の実施の形態との相異点を説明すると、CPU101は、後述の図13、図14のフローチャートに示す処理を実行する。音声素片データ115には、少なくとも声の性質が異なる2種類の音声素片データ（例えば、子供の声の音声素片データと、老人の声の音声素片データ）が含まれている。これらの音声素片データも、声の性質に対応した基本周波数を持つものである。そして、どちらかの声（例えば子供の声）が第1音声、他方の声（例えば老人の声）が第2音声として設定されているものとする。音声合成装置の他の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0113】

また、本発明の第3の実施の形態に係る音声合成装置は、上記第1の実施の形態と同様に、辞書114、音声素片データ115、メインルーチン初期化部201、音声処理初期化部202、通信データ処理部204、通信データ格納部206、表示文章データ格納部207、文章表示部208、音声波形生成部209（

音声波形生成手段)、音声出力部210(音声出力手段)、初期化部203・受信部205を有する通信処理部211、音声素片データ115、音響パラメータ212、出力パラメータ213を備えている(上記図2参照)。音声合成装置のプログラムモジュールの各部の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0114】

また、本発明の第3の実施の形態に係る音声合成装置の音声出力部210は、上記第1の実施の形態と同様に、一時蓄積部601、コントロール部602、音声再生部604、ミキシング部605を備えている(上記図3参照)。

【0115】

第3の実施の形態における上記第1の実施の形態との相異点を説明すると、音声再生部604は少なくとも2つ以上(実際は、同時に合成が予想される数だけ必要)存在し、コントロール部602は、音声波形603が送られてきた場合は、その時点で使用されていない音声再生部604に音声波形603を送って、再生を実行する。音声再生部604が出力する個々の音声データは、少なくとも2つ以上(実際は、同時に合成が予想される数だけ必要)の入力部を持つミキシング部605に送られ、ミキシング部605は音声データを合成して、最終的な合成音声データを上記図1に示したスピーカ112から出力する。

【0116】

また、コントロール部602は、音声波形生成部209から、どの音声で出力中かどうかの問い合わせを受け、使用中の全ての音声再生部604で再生中の音声波形のデータを調べて、音声波形生成部209に結果を返す機能も持っている。音声出力部210の他の構成は上記第1の実施の形態と同様であり、説明を省略する。

【0117】

次に、上記の如く構成された本発明の第3の実施の形態に係る音声合成装置の動作を図13、図14、図16を参照しながら詳細に説明する。下記の処理は上記図1に示したCPU101の制御のもとに実行される。

【0118】

図13は音声合成装置の通信データ処理部204から音声波形生成部209に送られてきたテキストデータを音声出力する処理を示す流れ図である。先ずステップS1301で、音声出力部210のコントロール部602に対して現在音声を出力中か問い合わせる。その結果、音声を出力中でなければ、ステップS1308にて、音声の種類を第1音声（例えば子供の声）に設定し、ステップS1304に進む。

【0119】

上記ステップS1301で現在音声を出力中であれば、ステップS1302にて、音声出力部210のコントロール部602に対して現在出力中の音声の種類を問い合わせ、現在出力中の音声に第1音声が含まれていないならば、ステップS1308にて、音声の種類を第1音声（例えば子供の声）に設定する。それ以外の場合は、ステップS1303にて、音声の種類を第2音声（例えば老人の声）に設定する。

【0120】

ステップS1304にて、上記ステップS1303もしくはステップS1308で切り替えた音声の種類の情報に合わせて、音声素片データ115の中から適切な種類の音声素片データが選択される。ステップS1305にて、辞書114を使ってテキストデータの言語解析を行い、テキストデータの読みと音調成分を生成する。更にステップS1306にて、音響パラメータ212に収納された予め設定された、声の高さ・アクセント・発声速度等に関するパラメータのうち、選択された音声の種類に対応したパラメータに従って、上記ステップS1304にて選択されていた音声素片データと、上記ステップS1305にて解析されたテキストデータの読みと音調成分を使って音声波形を生成する。

【0121】

ステップS1307にて、上記ステップS1306で生成した音声波形を音声出力部210に渡して音声出力を行う。音声波形が音声出力部210に送られるが、音声再生部604の1つを使って音声の再生が行われるが、音声再生部604にて現在再生中の音声がある場合は、ミキシング部605にて現在再生中の音声に新しく渡された音声がミキシングされて音声出力が行われることになる。現

在再生中の音声がない場合は、再生された音声はミキシング部605を通過するが、何の処理もなされずにそのままの音声出力が行われることになる。

【0122】

このように、複数の音声出力が重なり合うことを検知した時、それぞれの音声を違う種類の音声で出力することによって、複数の音声が重なっても容易に聞き取れるようになる。

【0123】

尚、3種類以上の音声が重なる可能性があるため、第3音声以降の音声も設定している場合は、図14に示すように、ステップS1403で、出力中でない優先順位の高い音声を選択すればよい（図14で、ステップS1403以外の部分は上記図13と全く同一の処理を行うので、繰り返しの説明は省略する）。

【0124】

図15は音声合成装置における第1音声での出力音声と第2音声での出力音声の時間的な関係を示す概念図、図16は音声合成装置における音声の種類を設定する方法を示す説明図である。

【0125】

キー ボード104またはPD105により、音声出力設定画面の指示があると、CPU101は描画部116を用いて図16に示す設定画面の画像データを生成し、表示コントローラ109によりモニタ110上に表示する。

【0126】

そして、ユーザは、PD105を用いて、図16の設定画面（設定手段）の503で、登録されている音声の中から第1音声とする音声を選択し、設定画面の504で、登録されている音声の中から第2音声とする音声を選択する。「OK」ボタン501を押下することにより、上記図1のRAM106上に格納されている第1音声、第2音声設定の変数が書き換えられ、選択が完了する。

【0127】

また、「cancel」ボタン502を押下したときは、RAM106上に格納されている第1音声、第2音声設定の変数が書き換えられず、選択は破棄されて音声の種類の設定モードは終了する。尚、第3音声以降の音声がある場合は、

上記503、504と同じ形式で第3音声等が選択できるようになっていればよい。

【0128】

以上説明したように、本発明の第3の実施の形態に係る音声合成装置によれば、複数の音声出力の重なり合いを検知し、それぞれの音声を違う種類の音声で出力することで、聞き取りが容易になるという効果を奏する。

【0129】

本実施形態を用いれば、例えばインターネットで接続された複数のユーザ端末同士がサーバコンピュータを介してテキストデータで会話をを行うチャットシステムにおいて、サーバコンピュータから送られてきた他のユーザの発言であるテキストデータを音声出力する際に、複数のユーザからのテキストデータが重なり合う場合に聞き取りを容易にすることができるという効果を奏する。

【0130】

〔第4の実施の形態〕

本発明の第4の実施の形態は、他のコンピュータ（サーバコンピュータ）から非同期に送られてくるテキストデータを音声出力するシステムにおいて、1つのテキストデータの音声出力が終了する前に、次のテキストデータが送られてきたとき、先に音声出力中の音声とは別の声の高さで次のテキストデータを読み上げるシステムである。

【0131】

本実施形態では、音声出力に重なりがないときに、通常に使用される音声を第1の高さの音声と呼び、音声が重なった場合に、次のテキストデータを読み上げるために使用される、先に音声出力中の第1の高さの音声とは異なる音声を第2の高さの音声と呼ぶ（図19参照）。本実施形態では、2つ以上の音声が重ならない前提で考えているが、更に音声が重なることが予想される場合は、第3の高さの音声、第4の高さの音声等を用意しておけばよい。

【0132】

本発明の第4の実施の形態に係る音声合成装置は、上記第3の実施の形態と同様に、CPU101、ハードディスクコントローラ（HDC）102、プログラ

ム113・辞書114・音声素片データ115を有するハードディスク(HD)103、キーボード104、ポインティングデバイス(PD)105、RAM106、通信回線インターフェイス(I/F)107、VRAM108、表示コントローラ109、モニタ110、サウンドカード111、スピーカ112を備えている(上記図1参照)。

【0133】

第4の実施の形態における上記第3の実施の形態との相異点を説明すると、CPU101は、後述の図17、図18のフローチャートに示す処理を実行する。音声合成装置の他の構成は上記第3の実施の形態と同様であり、説明を省略する。

【0134】

また、本発明の第4の実施の形態に係る音声合成装置は、上記第2の実施の形態と同様に、辞書114、音声素片データ115、メインルーチン初期化部201、音声処理初期化部202、通信データ処理部204、通信データ格納部206、表示文章データ格納部207、文章表示部208、音声波形生成部209(音声波形生成手段)、音声出力部210(音声出力手段)、初期化部203・受信部205を有する通信処理部211、音声素片データ115、音響パラメータ212、出力パラメータ213を備えている(上記図8参照)。音声合成装置のプログラムモジュールの各部の構成は上記第2の実施の形態と同様であり、説明を省略する。

【0135】

また、本発明の第4の実施の形態に係る音声合成装置の音声出力部210は、上記第3の実施の形態と同様に、一時蓄積部601、コントロール部602、音声再生部604、ミキシング部605を備えている(上記図3参照)。

【0136】

第4の実施の形態における上記第3の実施の形態との相異点を説明すると、音声再生部604は、コントロール部602の指示に従って再生時の声の高さを自由に調整できる機能を持っている。声の高さの調整は、例えば声を高い声にしたい場合は、再生される声の周波数成分のうち、高い音の周波数領域を強く出し、

その他の周波数領域を弱くすることによって可能となる。また、音声出力の重なりを検出し、それに対するアクションつまり声の高さを変更する制御はすべて音声出力部210で行われる。音声出力部210の他の構成は上記第3の実施の形態と同様であり、説明を省略する。

【0137】

次に、上記の如く構成された本発明の第4の実施の形態に係る音声合成装置の動作を図17、図18、図20を参照しながら詳細に説明する。下記の処理は上記図1に示したCPU101の制御のもとに実行される。

【0138】

図17は音声合成装置の音声波形生成部209より音声波形が音声出力部210に送られてきた時から音声を出力するまでの処理を示す流れ図である。先ずステップS1701で、音声出力部210のコントロール部602は、音声再生部604の動作状態を調べて、現在音声を出力中かどうかを確認する。その結果、音声を出力中でなければ、ステップS1708にて、音声を第1の高さの音声に設定し、ステップS1704に進む。

【0139】

上記ステップS1701で現在音声を出力中であれば、ステップS1702にて、コントロール部602は現在音声再生中の音声再生部604に対して現在出力中の音声の高さを問い合わせ、その結果、現在再生中の音声に第1の高さの音声が含まれていないならば、ステップS1708にて、音声を第1の高さの音声に設定する。それ以外の場合は、ステップS1703にて、音声を第2の高さの音声に設定する。

【0140】

ステップS1704では、音声再生部604の1つを使って音声波形の再生を行うのであるが、ここで上記ステップS1703もしくはステップS1708で設定した音声の高さの情報に合わせて、音声の高さを調整して再生を実行する。再生された音声は、ステップS1705で、音声のミキシングが行われ、最終的な音声の出力となる。この際、音声再生部604にて現在再生中の他の音声がある場合は、ミキシング部605で、現在再生中の音声に新しく再生された音声が

ミキシングされて音声出力が行われることになる。現在再生中の音声がない場合は、再生された音声はミキシング部605を通過するが、何の処理もなされずにそのままの音声出力が行われることになる。

【0141】

このように、複数の音声出力が重なり合うことを検知した時、それぞれの音声を違う高さの音声で出力することによって、複数の音声が重なっても容易に聞き取れるようになる。

【0142】

尚、3種類以上の音声が重なる可能性があるため、第3の高さ以降の音声も設定している場合は、図18に示すように、ステップS1803で、出力中でない優先順位の高い音声を選択すればよい（図18で、ステップS1803以外の部分は上記図17と全く同一の処理を行うので、繰り返しの説明は省略する）。

【0143】

図19は音声合成装置における第1の高さの音声での出力音声と第2の高さの音声での出力音声の時間的な関係を示す概念図、図20は音声合成装置における音声の高さを設定する方法を示す説明図である。

【0144】

キーボード104またはPD105により、音声出力設定画面の指示があると、CPU101は描画部116を用いて図20に示す設定画面の画像データを生成し、表示コントローラ109によりモニタ110上に表示する。

【0145】

そして、ユーザは、PD105を用いて、図20の設定画面（設定手段）の503で、登録されている音声の中から第1の高さの音声を選択し、設定画面の504で、登録されている音声の中から第2の高さの音声を選択する。「OK」ボタン501を押下することにより、上記図1のRAM106上に格納されている第1の高さの音声、第2の高さの音声の設定の変数が書き換えられ、選択が完了する。

【0146】

また、「cancel」ボタン502を押下したときは、RAM106上に格

納されている第1の高さの音声、第2の高さの音声の設定の変数が書き換えられず、選択は破棄されて音声の高さの設定モードは終了する。尚、第3の高さ以降の音声がある場合は、上記503、504と同じ形式で第3の高さの音声等が選択できるようになっていればよい。

【0147】

以上説明したように、本発明の第4の実施の形態に係る音声合成装置によれば、複数の音声出力の重なり合いを検知し、それぞれの音声を違う高さの音声で出力することで、聞き取りが容易になるという効果を奏する。

【0148】

本実施形態を用いれば、例えばインターネットで接続された複数のユーザ端末同士がサーバコンピュータを介してテキストデータで会話をを行うチャットシステムにおいて、サーバコンピュータから送られてきた他のユーザの発言であるテキストデータを音声出力する際に、複数のユーザからのテキストデータが重なり合う場合に聞き取りを容易にすることができるという効果を奏する。

【0149】

〔他の実施の形態〕

上述した本発明の実施の形態においては、インターネットを利用してテキストデータで会話をを行うチャットシステムに本発明を適用した例を上げたが、インターネット以外の他の通信媒体を利用してテキストデータで会話をを行うシステムにも本発明を適用可能である。

【0150】

尚、本発明は、複数の機器から構成されるシステムに適用しても、1つの機器からなる装置に適用してもよい。上述した実施形態の機能を実現するソフトウェアのプログラムコードを記憶した記憶媒体を、システム或いは装置に供給し、そのシステム或いは装置のコンピュータ（またはCPUやMPU）が記憶媒体等の媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0151】

この場合、記憶媒体等の媒体から読み出されたプログラムコード 자체が上述し

た実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体等の媒体は本発明を構成することになる。プログラムコードを供給するための記憶媒体等の媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROM、或いはネットワークを介してのダウンロードの方法などを用いることができる。

【0152】

また、コンピュータが読み出したプログラムコードを実行することにより、上述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOSなどが実際の処理の一部または全部を行い、その処理によって上述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0153】

更に、記憶媒体等の媒体から読み出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって上述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0154】

【発明の効果】

以上説明したように、請求項1～9記載の音声出力装置、請求項13～21記載の音声出力システム、請求項25～33記載の音声出力方法、請求項37～45記載の記憶媒体によれば、複数のテキストデータの合成音声を重ね合わせて发声する場合、複数のテキストデータをそれぞれ異なる種類の音声で音声合成し出力させるため、容易に複数のテキストデータの音声を聞き分けることができる音声出力装置を提供できるという効果を奏する。

【0155】

また、請求項10、11記載の音声出力装置、請求項22、23記載の音声出

力システム、請求項34、35記載の音声出力方法、請求項46、47記載の記憶媒体によれば、複数のテキストデータの合成音声を重ね合わせて発声する場合、複数のテキストデータの音声をそれぞれ異なる発声手段で発声させるため、容易に複数のテキストデータの音声を聞き分けることができる音声出力装置を提供できるという効果を奏する。

【0156】

また、請求項12記載の音声出力装置、請求項24記載の音声出力システム、請求項36記載の音声出力方法、請求項48記載の記憶媒体によれば、インターネットを介してテキストデータで会話をを行うシステムにおいても、上記と同様に、容易に複数のテキストデータの音声を聞き分けることができるという効果を奏する。

【図面の簡単な説明】

【図1】

本発明の第1、第3、第4の実施の形態に係る音声合成装置の概略構成を示すブロック図である。

【図2】

本発明の第1、第3の実施の形態に係る音声合成装置のプログラムのモジュール関係を示す説明図である。

【図3】

本発明の第1、第3、第4の実施の形態に係る音声合成装置のプログラムのモジュールにおける音声出力部の詳細構成を示す説明図である。

【図4】

本発明の第1の実施の形態に係る音声合成装置の音声波形生成部の処理を示す流れ図である。

【図5】

本発明の第1の実施の形態に係る音声合成装置における主性別での出力音声と副性別での出力音声の時間的な関係を示す概念図である。

【図6】

本発明の第1の実施の形態に係る音声合成装置の性別設定モード画面を示す説

明図である。

【図7】

本発明の第2の実施の形態に係る音声合成装置の概略構成を示すブロック図である。

【図8】

本発明の第2、第4の実施の形態に係る音声合成装置のプログラムのモジュール関係を示す説明図である。

【図9】

本発明の第2の実施の形態に係る音声合成装置のプログラムのモジュールにおける音声出力部の詳細構成を示す説明図である。

【図10】

本発明の第2の実施の形態に係る音声合成装置の音声出力部の処理を示す流れ図である。

【図11】

本発明の第2の実施の形態に係る音声合成装置における両方のスピーカでの再生音声と別々のスピーカでの再生音声の時間的な関係を示す概念図である。

【図12】

本発明の第2の実施の形態に係る音声合成装置のスピーカ設定モード画面を示す説明図である。

【図13】

本発明の第3の実施の形態に係る音声合成装置の音声波形生成部の処理を示す流れ図である。

【図14】

本発明の第3の実施の形態に係る音声合成装置の音声波形生成部の処理を示す流れ図である。

【図15】

本発明の第3の実施の形態に係る音声合成装置における第1音声での出力音声と第2音声での出力音声の時間的な関係を示す概念図である。

【図16】

本発明の第3の実施の形態に係る音声合成装置の音声種類設定モード画面を示す説明図である。

【図17】

本発明の第4の実施の形態に係る音声合成装置の音声出力部の処理を示す流れ図である。

【図18】

本発明の第4の実施の形態に係る音声合成装置の音声出力部の処理を示す流れ図である。

【図19】

本発明の第4の実施の形態に係る音声合成装置における第1の高さの音声での出力音声と第2の高さの音声での出力音声の時間的な関係を示す概念図である。

【図20】

本発明の第4の実施の形態に係る音声合成装置の声の高さ設定モード画面を示す説明図である。

【図21】

本発明の音声出力方法を実行するプログラム及び関連データを記憶した記憶媒体の記憶内容の構成例を示す説明図である。

【図22】

本発明の音声出力方法を実行するプログラム及び関連データが記憶媒体から装置に供給される概念例を示す説明図である。

【符号の説明】

101 CPU

110 モニタ

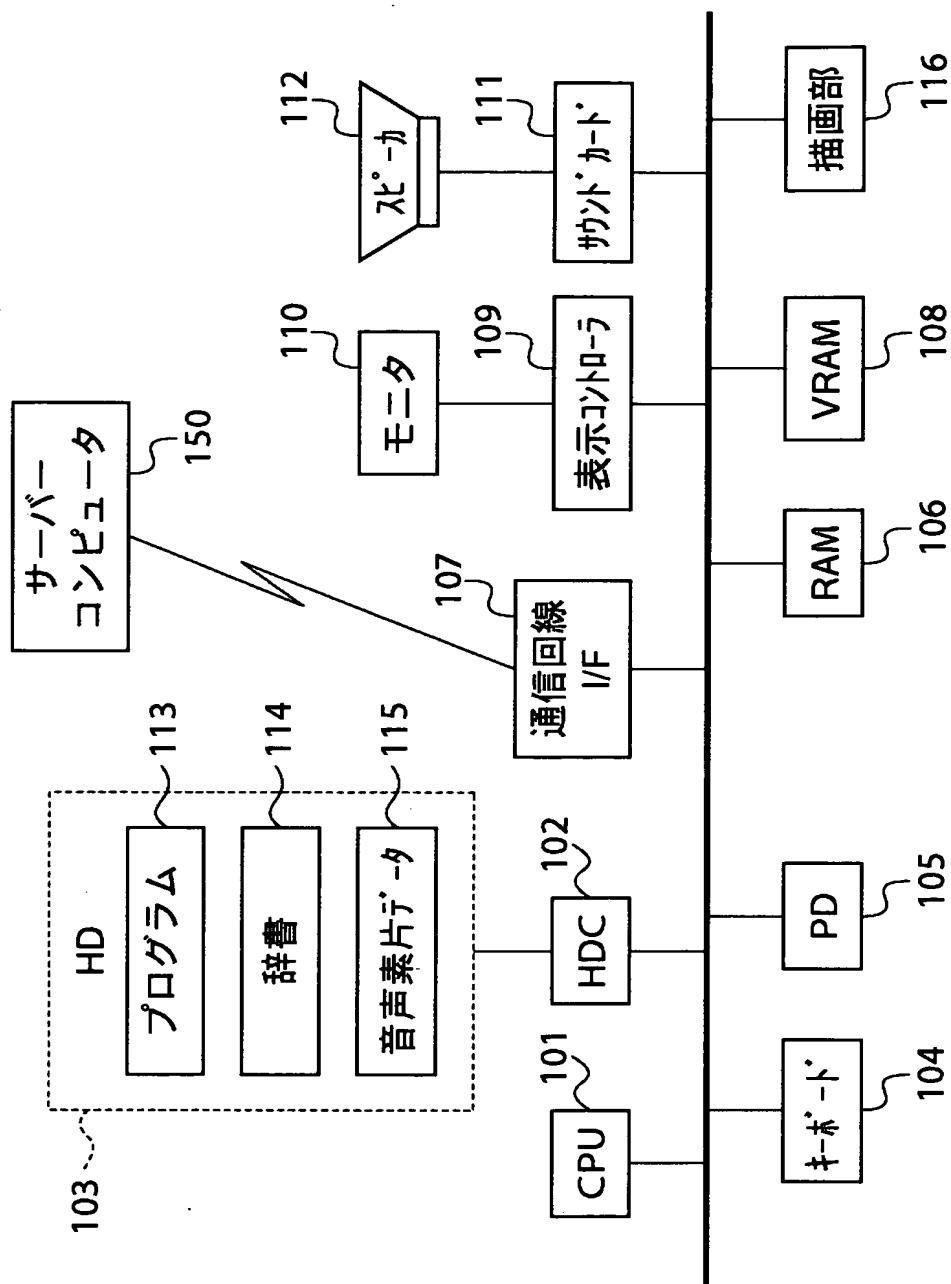
112R、112L スピーカ

209 音声波形生成部

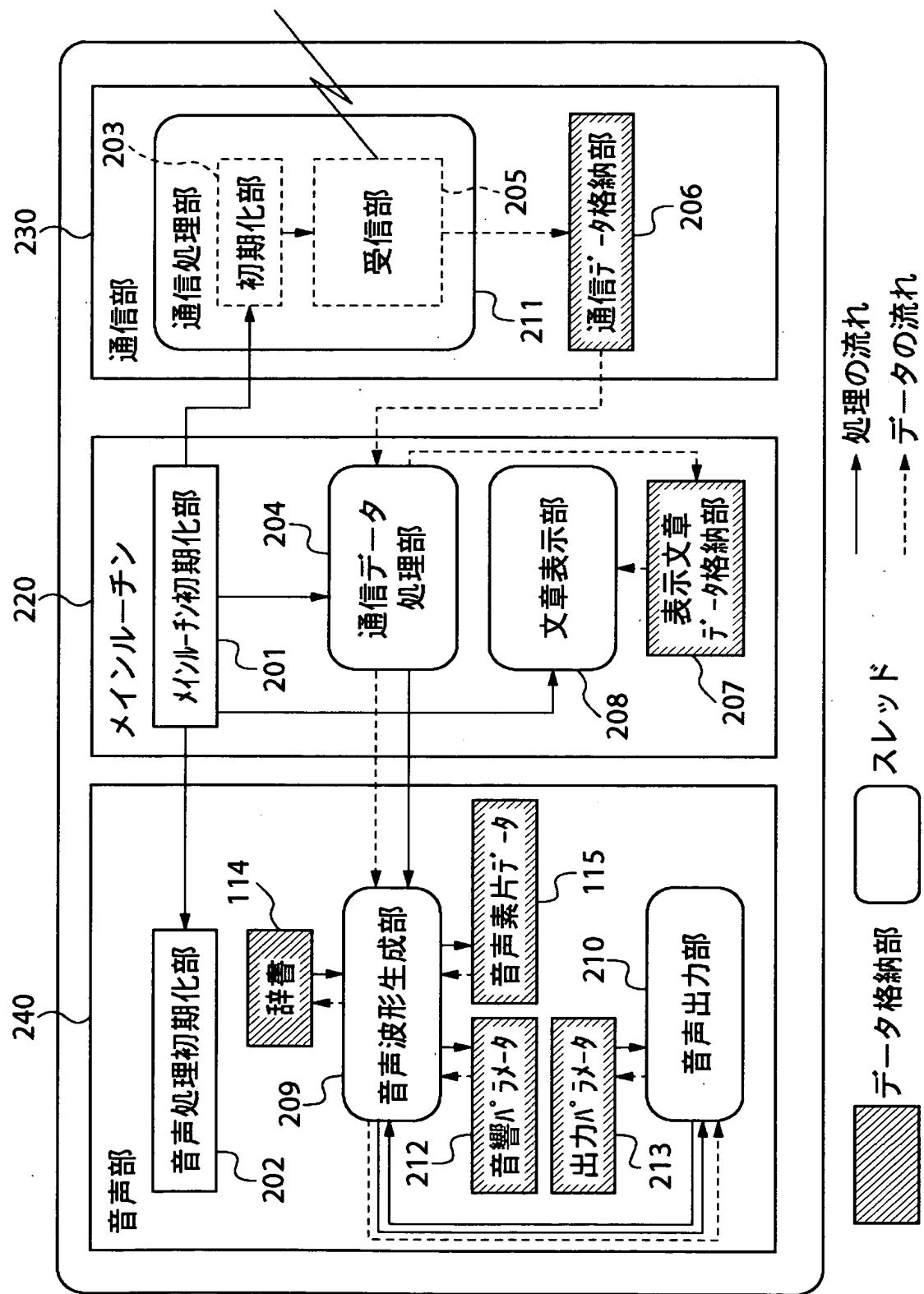
210 音声出力部

【書類名】図面

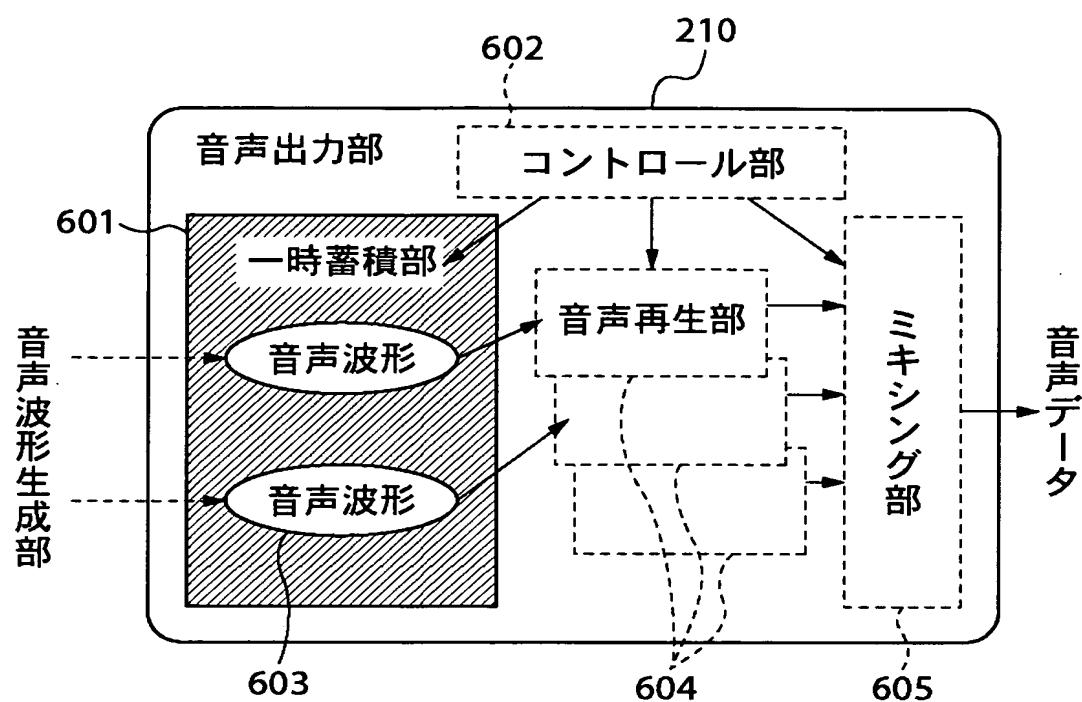
【図1】



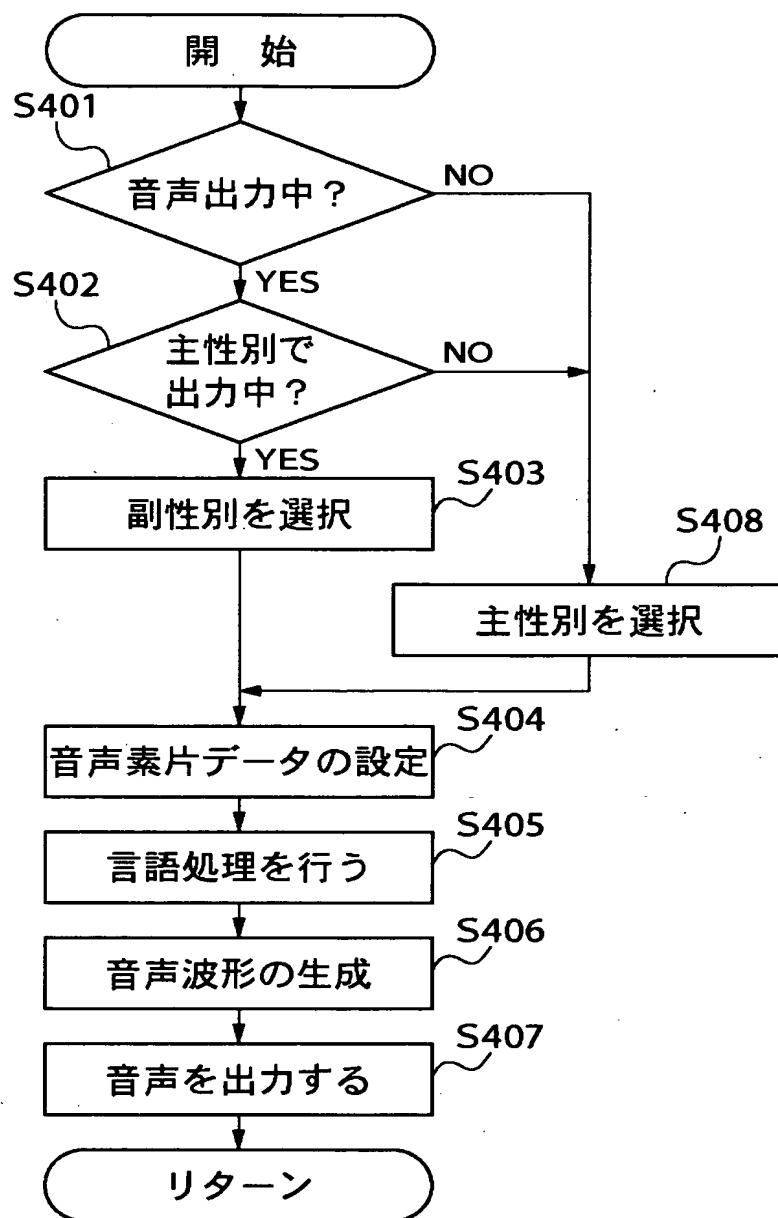
【図2】



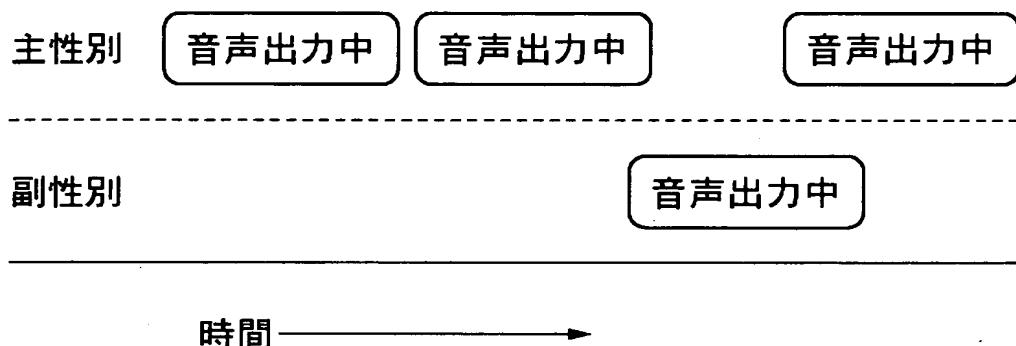
【図3】



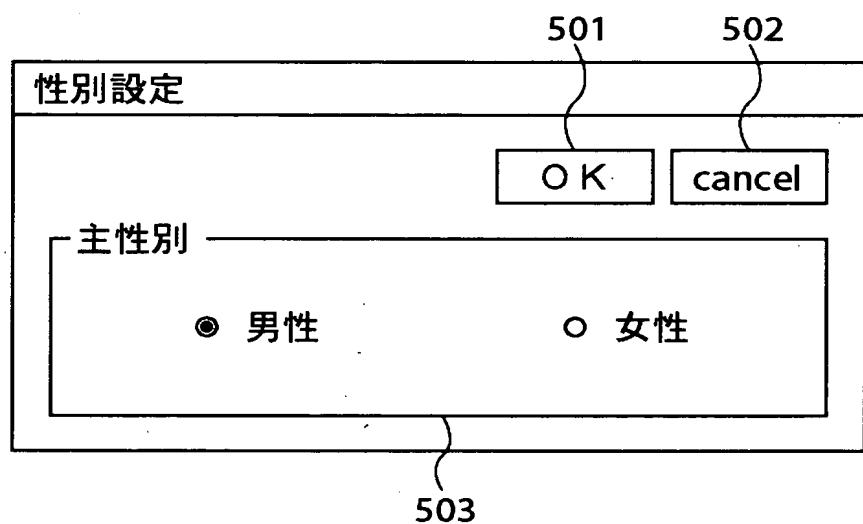
【図4】



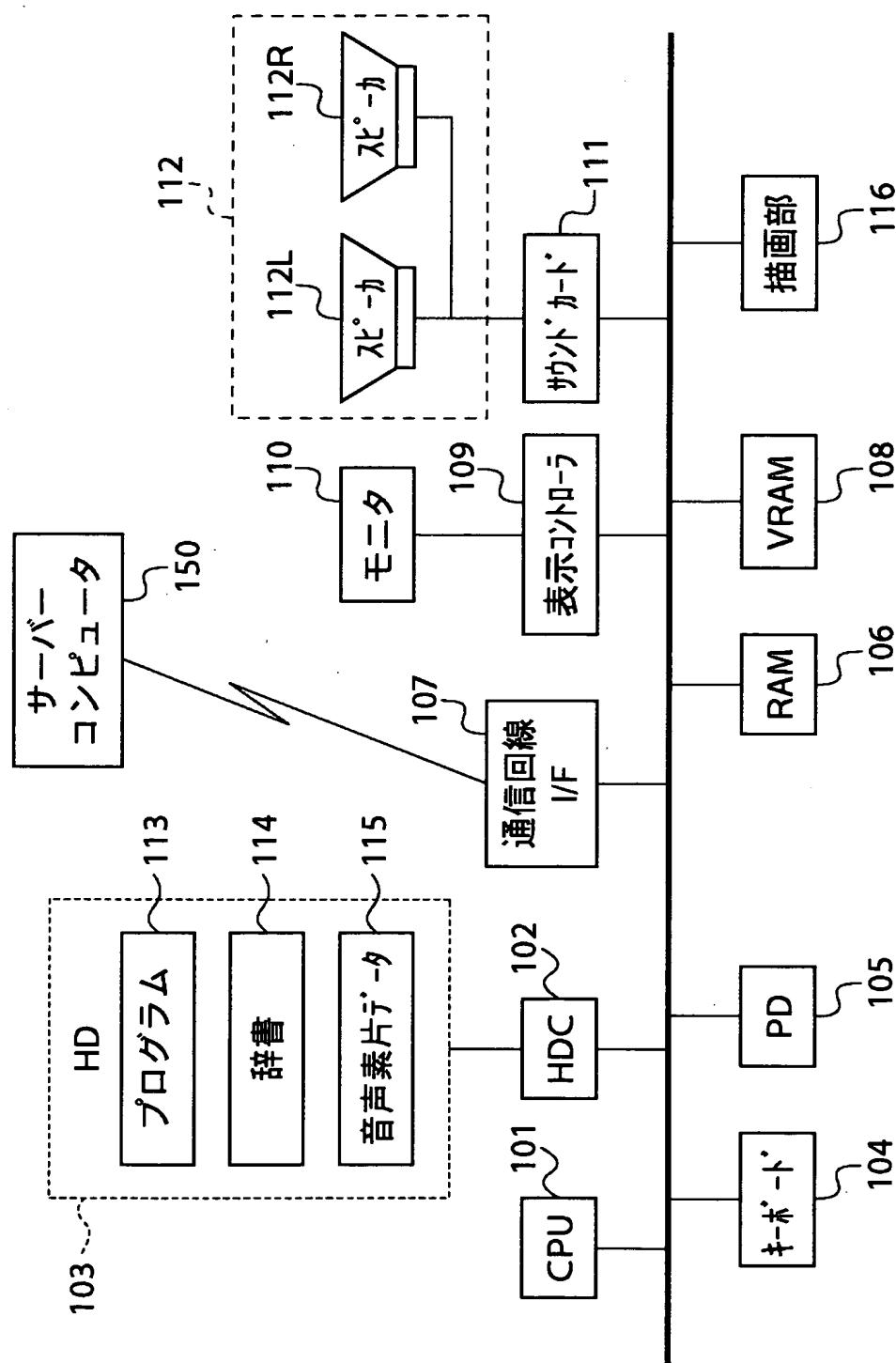
【図5】



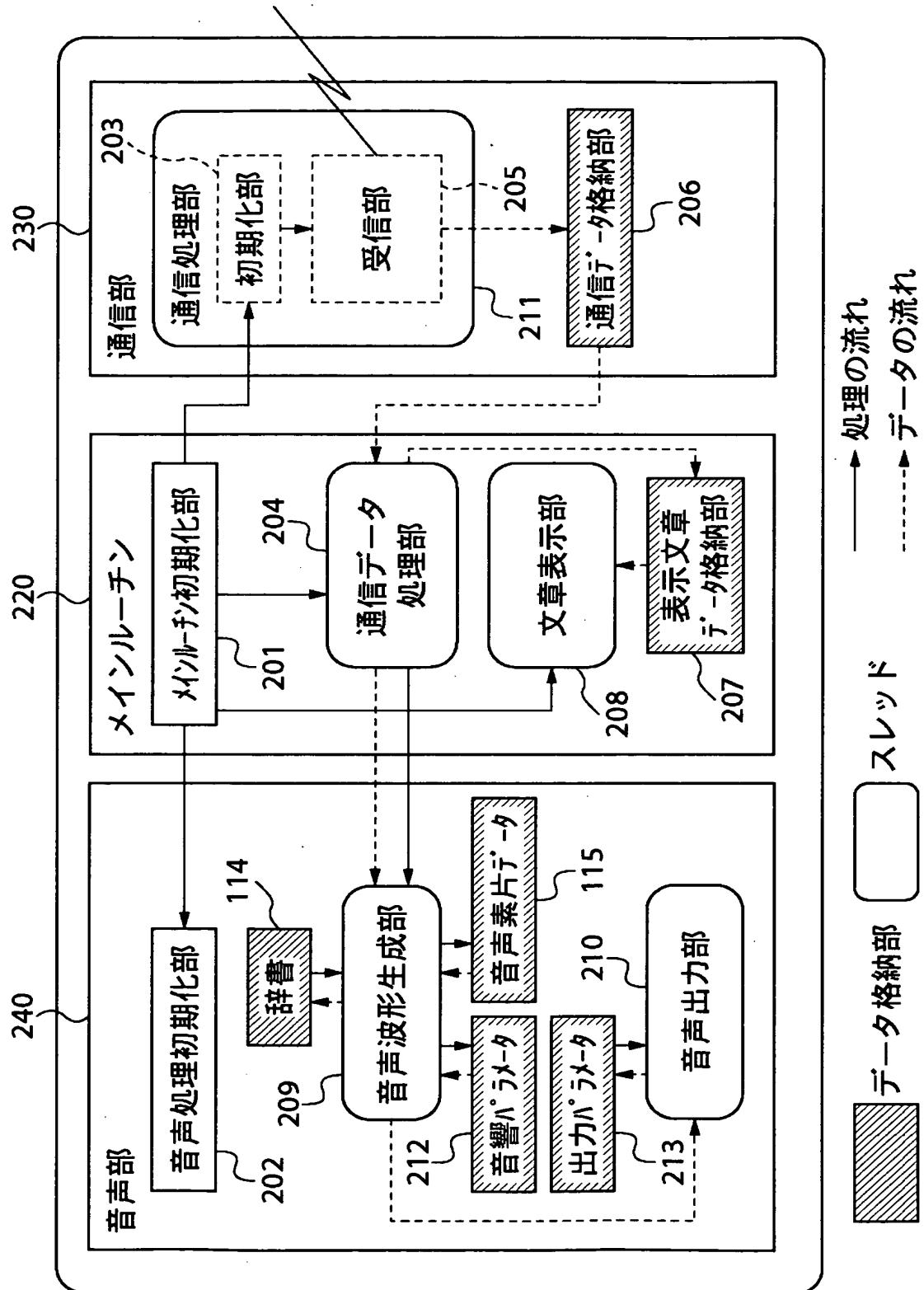
【図6】



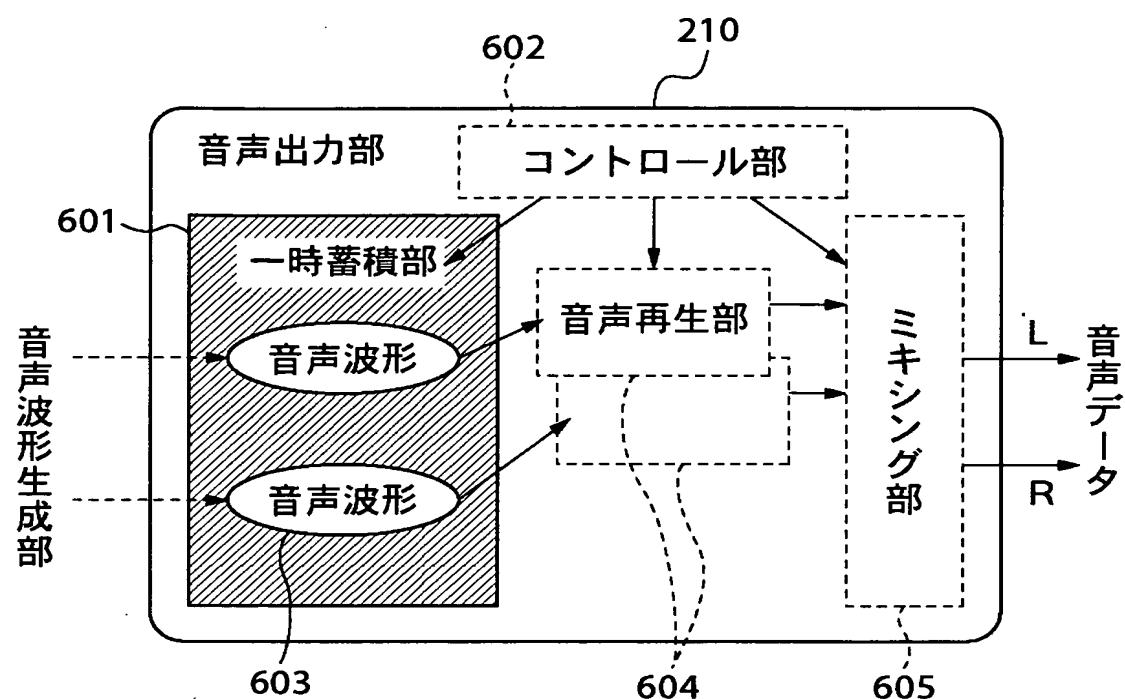
【図7】



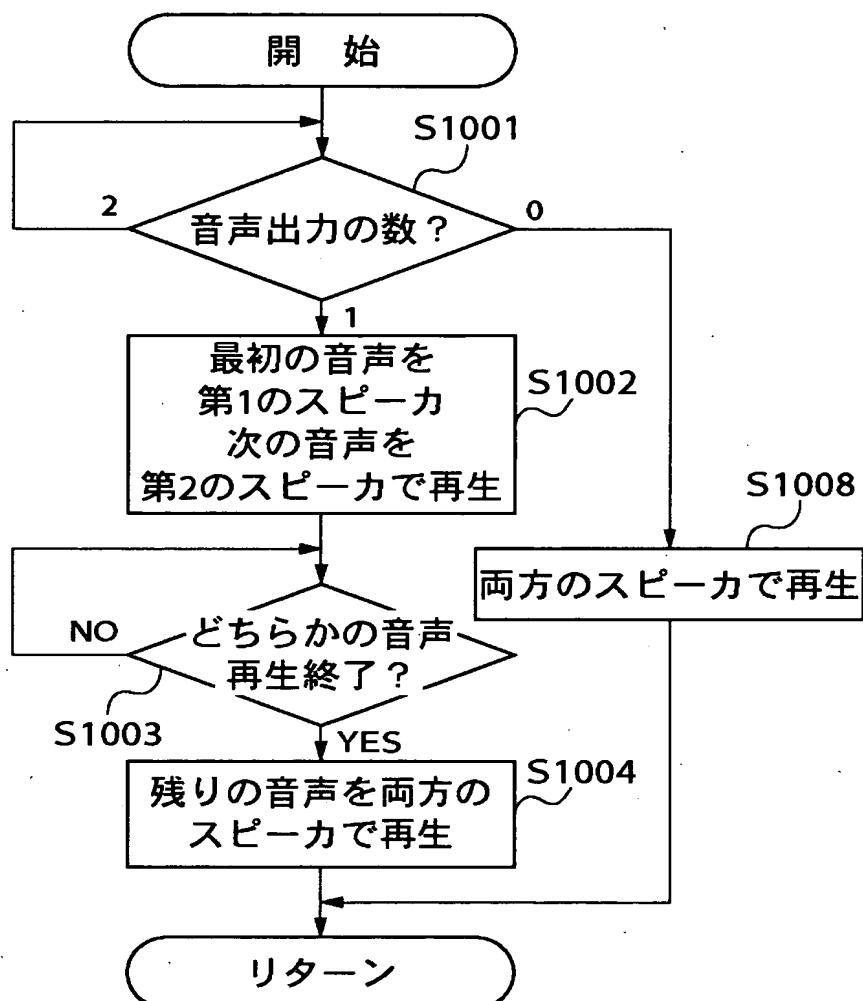
〔図8〕



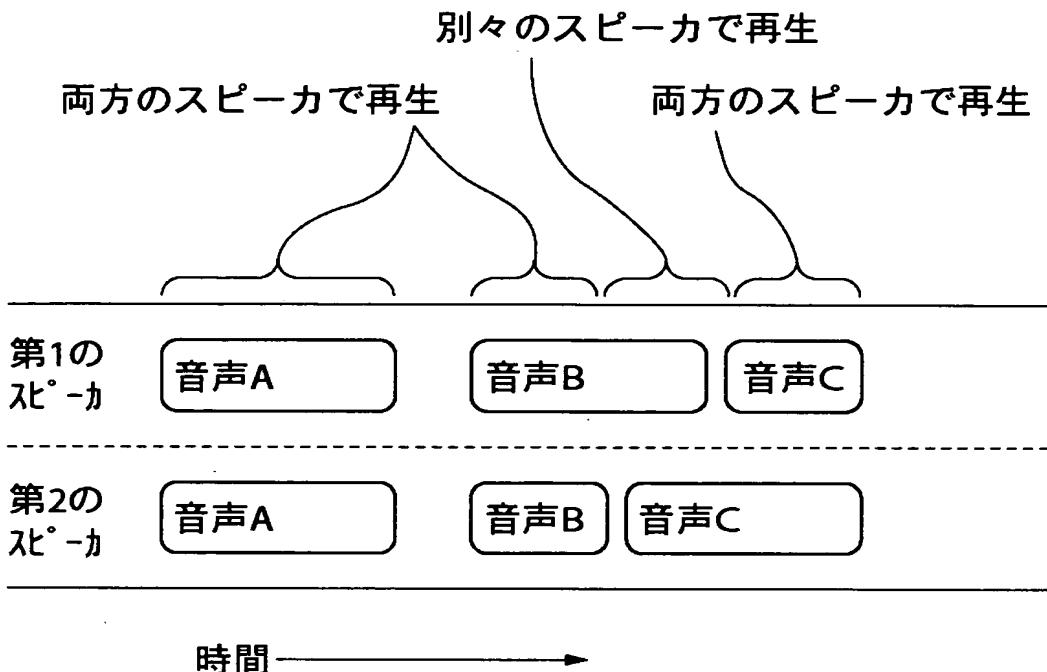
【図9】



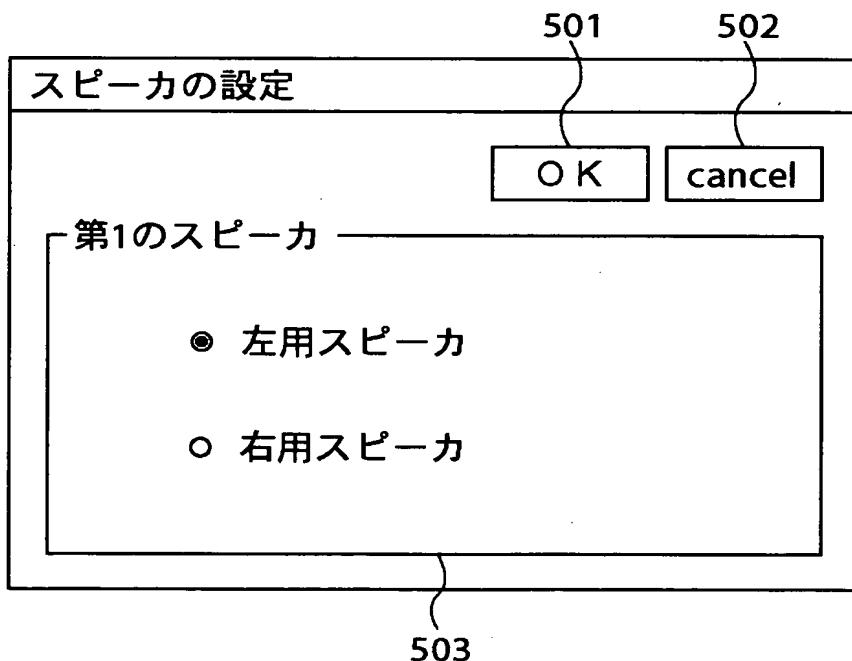
【図10】



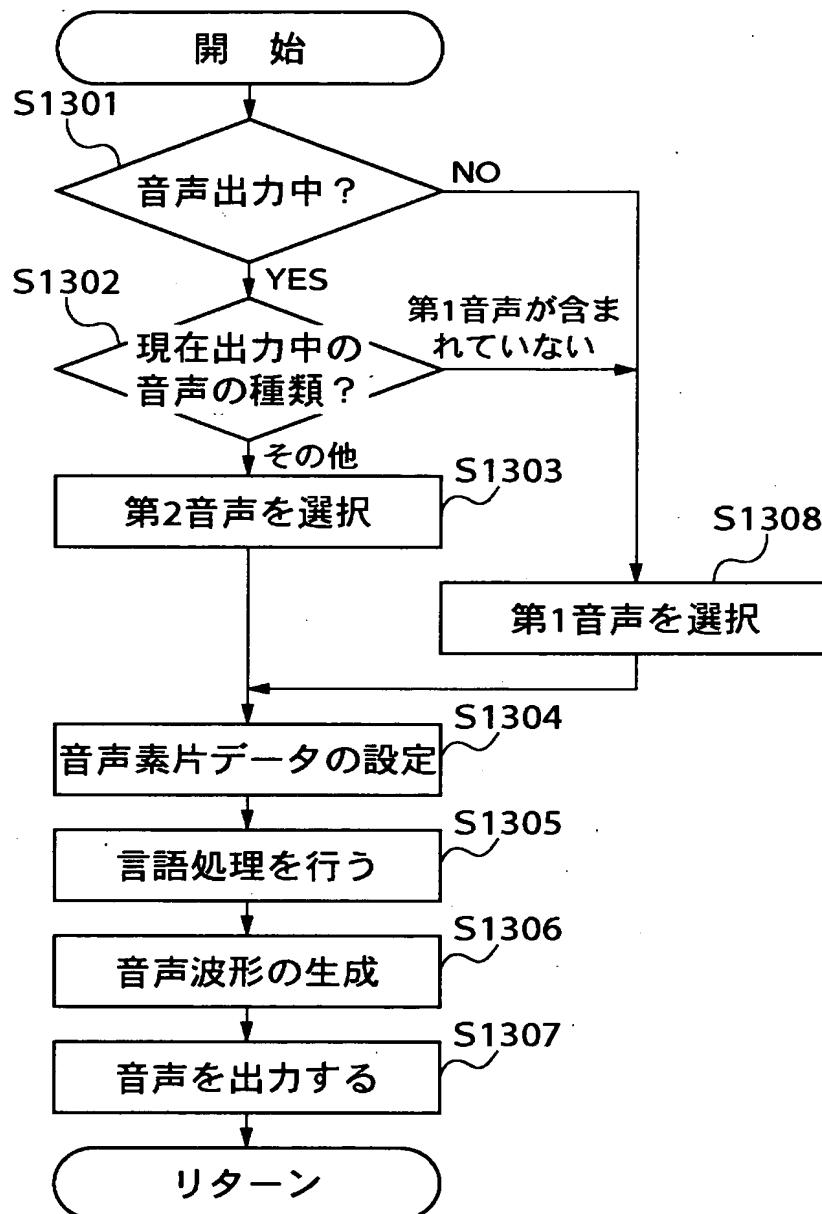
【図11】



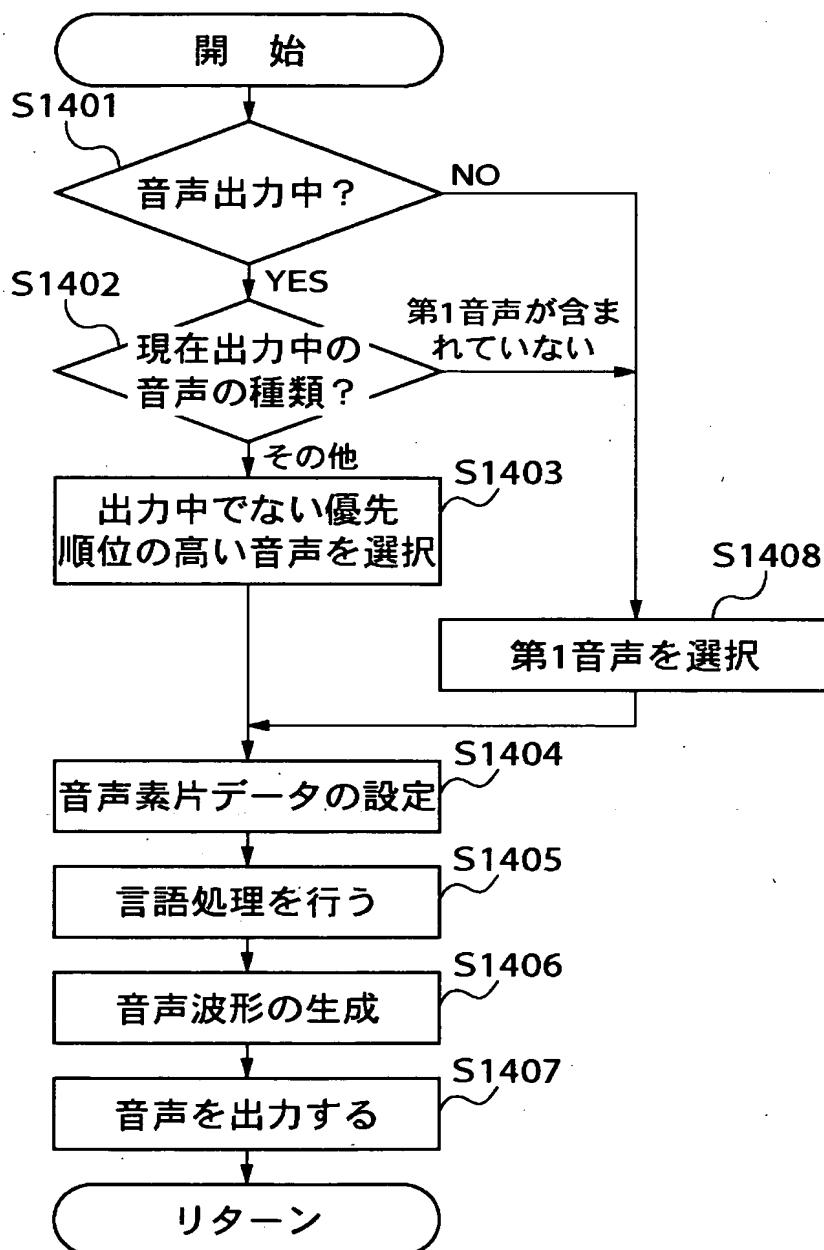
【図12】



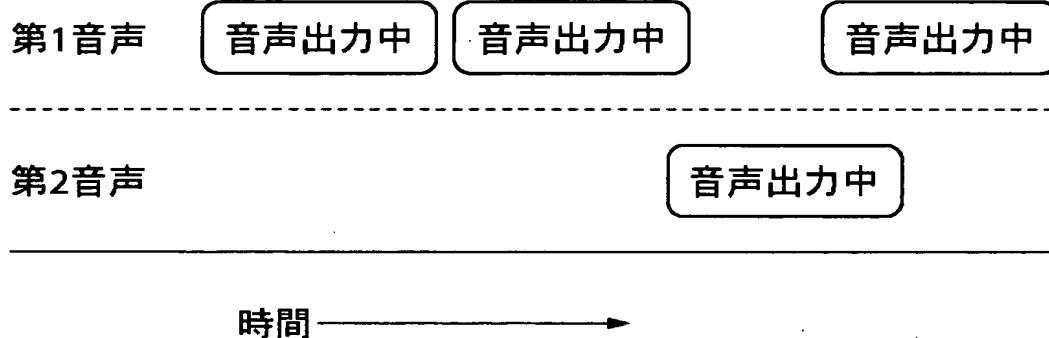
【図 1-3】



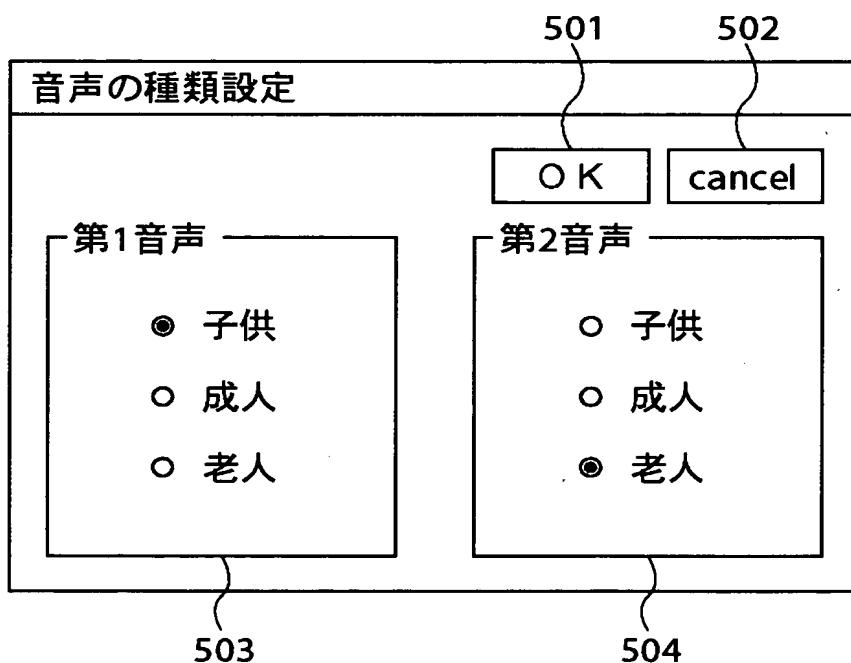
【図14】



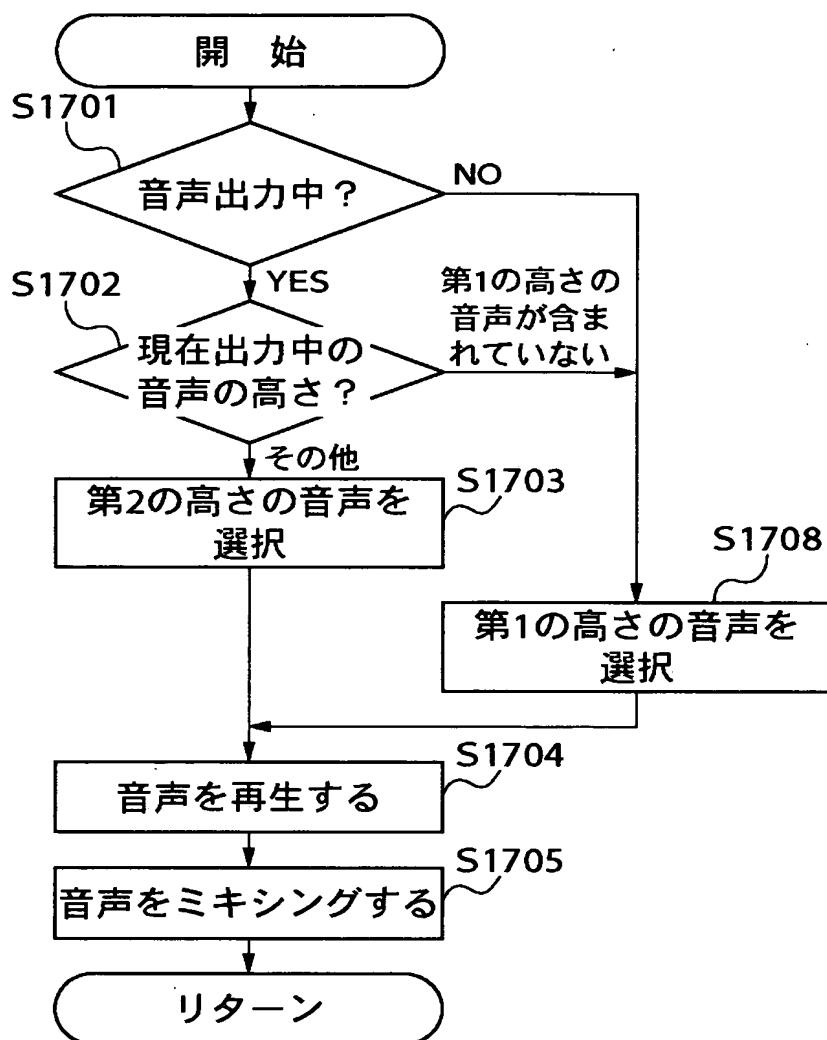
【図15】



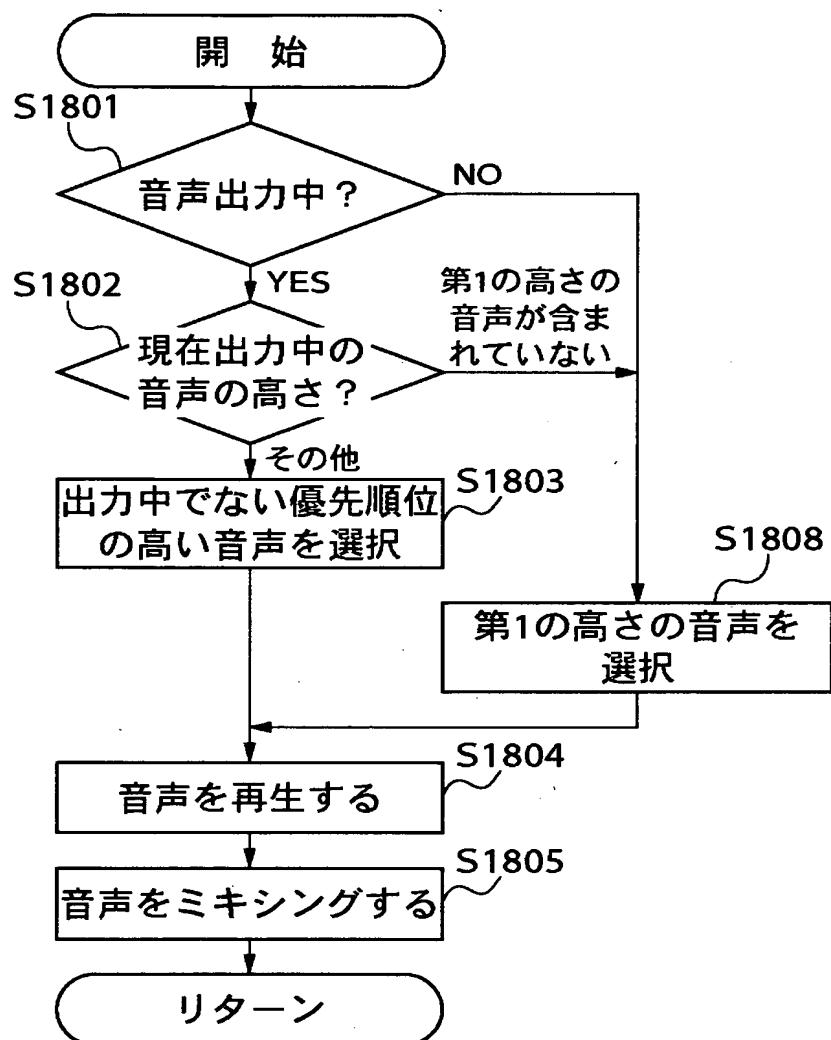
【図16】



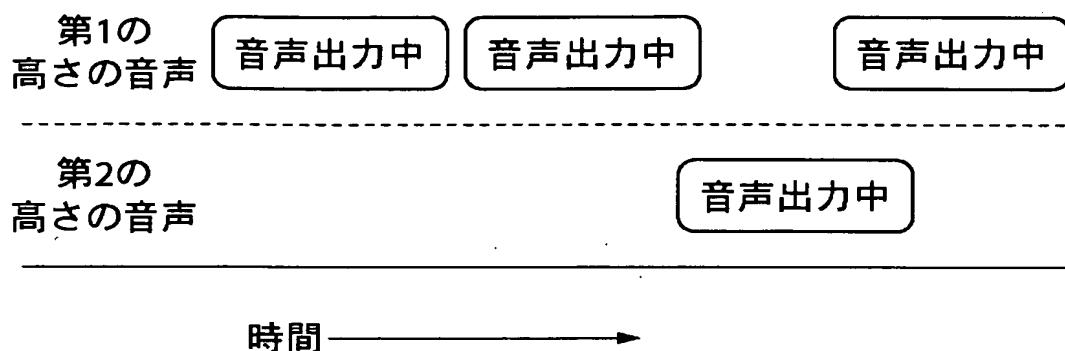
【図17】



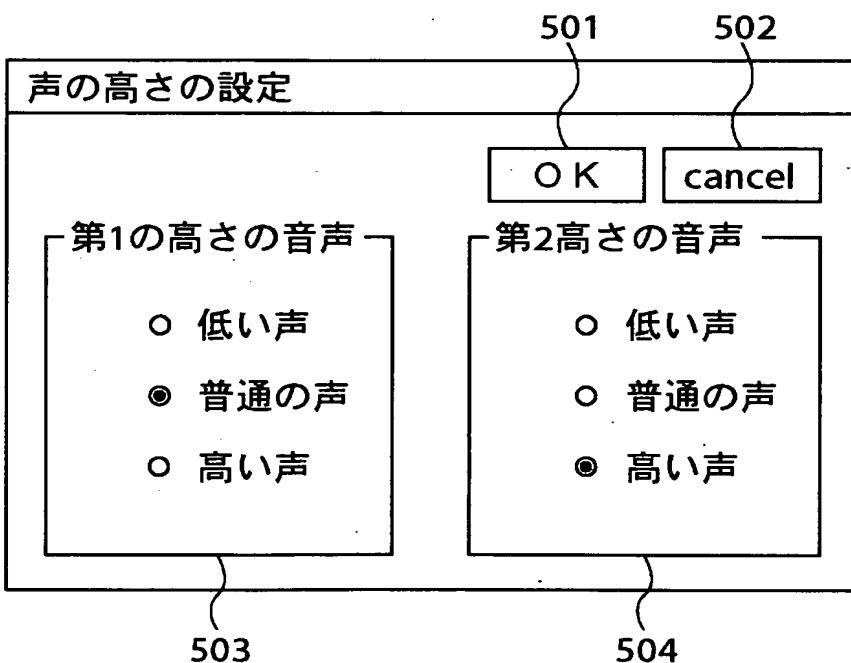
【図18】



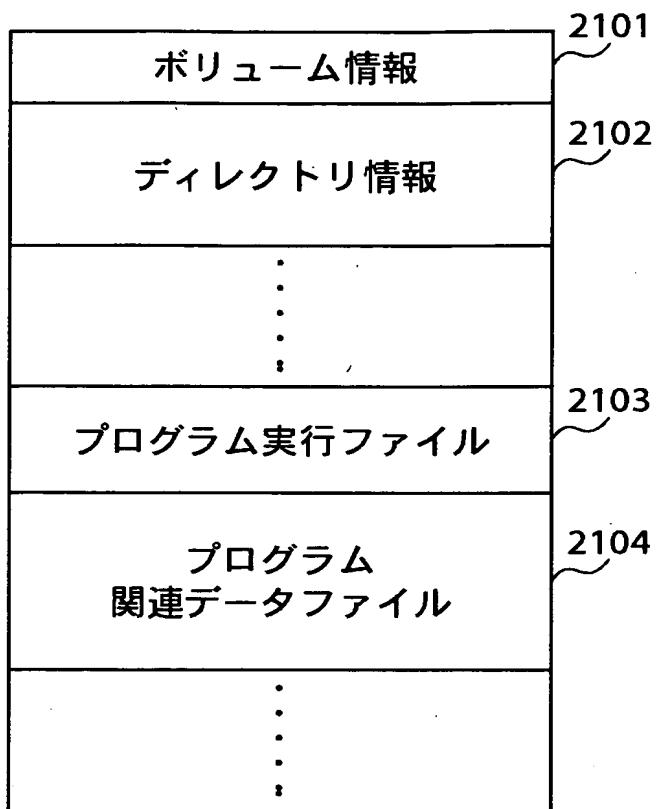
【図19】



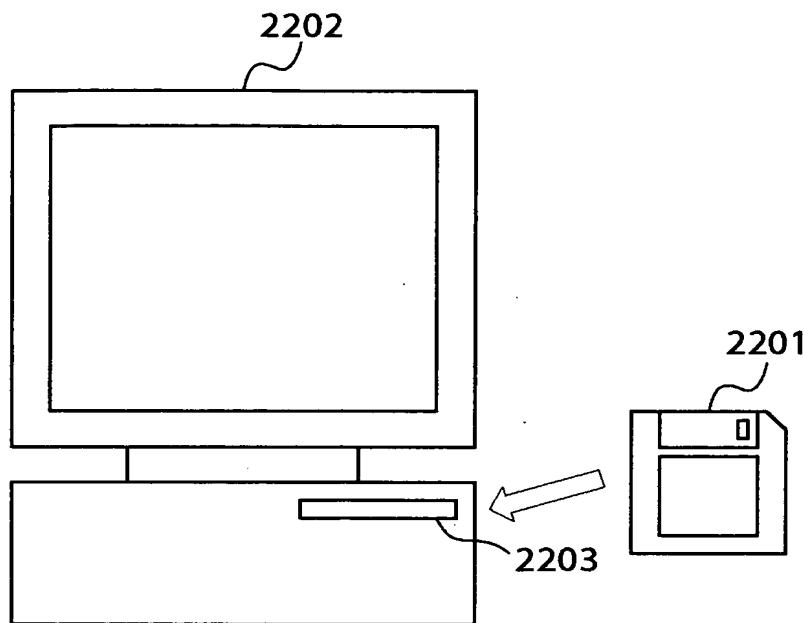
【図20】



【図21】



【図22】



【書類名】 要約書

【要約】

【課題】 複数のテキストデータの合成音声を重ね合わせて発声する場合、複数のテキストデータをそれぞれ異なる種類の音声で音声合成し出力させることで、容易に複数のテキストデータの音声を聞き分けることができる音声出力装置、音声出力システム、音声出力方法及び記憶媒体を提供する。

【解決手段】 テキストデータの音声波形を生成する音声波形生成部209と、複数のテキストデータの音声出力が重なり合うことを検知した場合は、それぞれのテキストデータを違う性別（主性別、副性別）の音声で、或いは別のスピーカから、或いは違う種類の音声で、或いは違う高さの音声で出力させる音声出力部210とを具備する。

【選択図】 図2

出願人履歴情報

識別番号 [000001007]

1. 変更年月日 1990年 8月30日

[変更理由] 新規登録

住 所 東京都大田区下丸子3丁目30番2号

氏 名 キヤノン株式会社